IARC HANDBOOKS OF CANCER PREVENTION

International Agency for Research on Cancer
World Health Organization

Volume 12

Methods for Evaluating Tobacco Control Policies

IARC
2008

IARC HANDBOOKS OF CANCER PREVENTION

Tobacco Control

IARC

International Agency for Research on Cancer
World Health Organization

Volume 12

Methods for Evaluating Tobacco Control Policies

IARC
2008

This Handbook was made possible thanks to the generous funding by the Ministère de la Santé, de la Jeunesse et des Sports

International Agency for Research on Cancer

The International Agency for Research on Cancer (IARC) was established in 1965 by the World Health Assembly, as an indepently financed organization within the framework of the World Health Organization. The headquarters of the Agency are in Lyon, France.

The Agency conducts a programme of research concentrating particularly on the epidemiology of cancer and the study of potential carcinogens in the human environment. Its field studies are supplemented by biological and chemical research carried out in the Agency's laboratories in Lyon and, through collaborative research agreements, in national research institutions in many countries. The Agency also conducts a programme for the education and training of personnel for cancer research.

The publications of the Agency contribute to the dissemination of authoratative information on different aspects of cancer research. Information about IARC publications, and how to order them, is available via the Internet at: http://www.iarc.fr/.

This publication represents the views and opinions of an IARC Working Group on Methods for Evaluating Tobacco Control Policies which met in Lyon, France, 12 - 19 March 2007.

Funding

The IARC Tobacco Control Handbook Volume 12 was funded by the Ministère de la Santé, de la Jeunesse et des Sports, France.

Published by the International Agency for Research on Cancer,
150 cours Albert Thomas, 69372 Lyon Cedex 08, France

Distributed by
WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel: +41 22 791 3264;
fax: +41 22 791 4857; email: bookorders@who.int).

Format for bibliographic citation:

IARC Handbooks of Cancer Prevention, Tobacco Control, Vol. 12: Methods for Evaluating Tobacco Control Policies (2008: Lyon, France).

IARC Library Cataloguing in Publication Data

Methods for evaluating tobacco control policies / IARC Working Group on Methods
for Evaluating Tobacco Control Policies, Lyon, France, International Agency for Research on Cancer, 2008

(IARC Handbooks of Cancer Prevention ; 12)

1. Health Policy 2. Neoplasms - prevention & control 3. Program Evluation
4. Smoking - prevention & control 5. Tobacco Industry - legislation and jurisprudence.
I. IARC Working Group on Methods for Evaluating Tobacco Control Policies II. Series

ISBN 978-92-832-3012-0 (NLM Classification QZ 39)
ISSN 1027-5622

Table of Contents

LIST OF PARTICIPANTS

Ron Borland *(Co-Chair)*
Cancer Control Research Institute
The Cancer Council Victoria
1 Rathdowne Street
Carlton, Victoria 3053
Australia

K. Michael Cummings *(Co-Chair)*
Department of Health Behaviour
Roswell Park Cancer Institute
Elm and Carlton Streets
Buffalo, NY 14263
USA

Timothy Baker *(not attending)*
Center for Tobacco Research and Intervention
Univeristy of Wisconsin Medical School
1930 Monroe Street, Suite 200
Madison, WI 53711-2027
USA

Ursula Bauer
Tobacco Control Program
New York State Department of Health
ESP Corning Tower, Room 710
Albany, NY 12237-0676
USA

Frank J. Chaloupka
Economics, College of Business Administration
Health Policy and Administration
University of Illinois at Chicago
601 S. Morgan St, Room 2103
Chicago, IL 60607-7121
USA

Carolyn Dresler *(not attending)*
Tobacco Prevention and Cessation Program
Arkansas Department of Health
4815 W Markham St.
PO Box 1437, Slot H-3
Little Rock, AR 72203-1437
USA

Jean-Francois Etter
Faculte de Medecine
Universite de Geneve
1 rue Michel-Servet
CH-1211 Geneve 4
Switzerland

Geoffrey T. Fong
Ontario Institute for Cancer Research and Department of Psychology
University of Waterloo
200 University Avenue West
Waterloo, Ontario N2L 3G1
Canada

Gary A. Giovino
Department of Health Behavior
School of Public Health and Health Professions
SUNY at Buffalo
622 kimball Tower
Buffalo, NY 14214-3079
USA

G. Emmanuel Guindon
Centre for Health Economics and Policy Analysis
Health Sciences Centre 3H1 area
McMaster University
1200 Main Street West
Hamilton, Ontario L8N 3Z5
Canada

Prakash C. Gupta
Healis-Sekhsaria Inst. for Public Health
Plot No. 28, Sector 11
CBD Belalpur
601/B Great Eastern Chambers
Navi Mumbai
India

David Hammond
Department of Health Studies and Gerontology
University of Waterloo
200 University Avenue West
Waterloo, Ontario N2L 3G1
Canada

Gerard Hastings *(not attending)*
Centre for Tobacco Control Research
University of Stirling and the Open University
Stirling FK9 4LA
Scotland

Andrew Hyland
Department of Health Behaviour
Roswell Park Cancer Institute
Elm and Carlton Streets
Buffalo, NY 14263
USA

Luk Joossens, *(not attending)*
Belgian Foundation Against Cancer
479 Chaussée de Louvain
B-1030 Brussels
Belgium

Alan Lopez, *(not attending)*
The University of Queensland
Herston Road
Herston Qld 4006
Australia

Anne Marie MacKintosh *(not attending)*
Institute for Social Marketing
University of Stirling and the Open University
Stirling FK9 4LA
Scotland

Ann McNeill
Division of Epidemiology & Public Health
University of Nottingham
Clinical Sciences Building
Notthingham NG5 1BP
UK

Mark Parascandola
Tobacco Control Research Branch
National Cancer Institute
6130 Executive Blvd. MSC 7337
Bethesda, MD 20892
USA

Armando Peruga
Tobacco Free Initiative
World Health Organization
Geneva
Switzerland

Patrick Petit
Tobacco Free Initiative
World Health Organization
Geneva
Switzerland

Megan E. Piper
Center for Tobacco Research & Intervention
University of Wisconsin
Medical School
1930 Monroe St., Suite 200
Madison, WI 53711-2027
USA

James F. Thrasher
Health Promotion, Education and Behavior
School of Public Health
University of South Carolina
800 Sumter Street, Room # 215
Columbia, SC 29208
USA; and
Instituto Nacional de Salud Pública,
Cuernavaca,
Mexico

Charles (Wick) Warren
Office on Smoking and Health
Centers for Disease Control and Prevention
4770 Buford Highway, NE
Atlanta, GA 30341-3717
USA

Representatives
Nathan Jones
Office on Smoking and Health
Global Tobacco Control Program
Centers for Disease Control and Prevention
4770 Buford Highway, NE
Atlanta, GA 30341-3717
USA

Martina Potschke-Langer
Cancer Prevention and WHO Collaborating Center for Tobacco Control
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 280 D-69120 Heidelberg
Germany

IARC Secretariat
Andrea Altieri
Robert Baan
Julien Berthiller
Paolo Boffetta (Group Head)
Lars Egevad
Fabrizio Giannandrea (Post-Meeting)
Julia Heck
María E. León (Responsible Officer)
Beatrice Secretan
Kurt Straif

Administrative assistance
Catherine Benard (Secretarial)
Latifa Bouanzi (Library)
John Daniel (Editor)
Jennifer Donaldson (Editor)
Roland Dray (Graphics)
Sharon Grant (Library)
Georges Mollon (Photography)
Sylvia Moutinho (Secretarial)
Annick Rivoire (Secretarial)
Josephine Thevenoux (Layout)

Acknowledgements

The Working Group acknowledges the major contribution to the work presented in this Handbook by Mary E. Thompson (University of Waterloo, Waterloo, Ontario, Canada), Daniel M. Bolt (University of Wisconsin, Madison, Wisconsin, USA), Matthew C. Farrelly (RTI International, Research Triangle Park, North Carolina, USA), Timothy P. Johnson (University of Illinois, Chicago, Illinois, USA) and Karl E. Wende and Jennifer L. Graf (University of Buffalo, Buffalo, New York, USA).

The IARC secretariat is grateful to the staff of the Libraries at the International Agency for Research on Cancer, Lyon, France and the World Health Organization, Geneva, Switzerland.

Preface

The *IARC Handbooks on Cancer Prevention* have traditionally presented the scientific evidence on the effects of interventions, such as sun protection or dietary chemoprevention, on preventing cancer, as well as the evaluation of the strength of the evidence in addressing the alleged protective effect.

In Volume 11, the first dedicated to tobacco control, the effects of smoking cessation on the risk of developing or dying of cancer, cardiovascular diseases, or chronic obstructive pulmonary disease were examined. In that volume, the health benefits of quitting smoking were investigated by comparing epidemiological studies reporting the risk of disease in never, former, and current smokers, as well as differences in risk with length of smoking abstinence, when available. An evaluation of the weight of the evidence was given for each disease contemplated.

For IARC, Volume 11 was exceptional in including disease outcomes other than cancer. Given the prominent etiologic position of smoking in other disease outcomes, limiting the review to cancer would have given a partial picture of the benefits derived from quitting smoking. How individuals overcome the smoking habit to achieve sustained abstinence has not been covered in the Handbooks. However, we know from numerous publications that one way of inducing quitting in a proportion of the population of smokers is through policy measures, implemented by local, regional, and/or national governments, intended to reduce both the number of smokers and the amount smoked in persistent users (e.g. by increasing the cost of tobacco products through the use of pricing and taxation policies). Interventions, which have been implemented at the individual and societal level to control the use of tobacco and concomitant health effects, have been adopted at different paces and with varying degrees of comprehensiveness in countries around the world, generating an irregular response to the tobacco epidemic. These interventions have included, to list a few, total or partial bans on smoking in work and public places; suppression of tobacco advertising, promotion, and sponsorship; anti-tobacco education and communication campaigns to raise awareness; changes to tobacco product labeling; and smoking cessation services.

A global, coordinated effort to use legislation and associated programmes to arrest the tobacco use epidemic is now led by the World Health Organization through the Framework Convention on Tobacco Control (WHO FCTC). The WHO FCTC encompasses a range of measures, in their totality representing a comprehensive approach designed to control tobacco use and supply. The body of policies stipulated in the WHO FCTC treaty became binding international law on February 27, 2005. Of the 38 articles, articles 6 to 14 cover policy interventions directed at preventing tobacco use, decreasing consumption, reducing toxicity, protecting non-smokers, and diminishing tobacco use initiation. Articles 15 to 17 relate to measures controlling the availability of tobacco (WHO, 2003). In other words, the policies are a series of measures conceived to counteract multiple domains of tobacco availability and use. The joint observance of the treaty by countries around the world will make it a global response to the tobacco epidemic. However, the reach of the policy interventions included in the WHO FCTC will depend on how effectively countries formulate and implement these policies. As of November 7, 2008, 161 countries have become parties to the treaty (http://www.who.int/tobacco/framework/en/index.html; accessed November 10, 2008).

The FCTC has propelled tobacco control into a new era, as

countries all over the world incorporate its policies and recommendations into their own laws. As tobacco control policies are formulated and implemented, it is important that they undergo rigorous evaluation. In the same way that evidence-based medicine has been built from thorough evaluation of treatment options, evidence-based public health must build on a database of rigorous evaluations of public health policies. Such knowledge will allow implementation of the most powerful policy interventions, and will do so in ways that will maximize their effectiveness.

Towards this goal, IARC convened a working group of international tobacco control experts from March 12-19, 2007 to propose a framework for guiding the evaluation of tobacco control policies expected to be formulated worldwide in response to WHO FCTC. Four broad questions were considered by the working group, each with several more specific related sub-questions, to guide the review of the scientific literature on the methods and measures of tobacco policy evaluation. The broad questions cover how the effects of a policy are determined, the core constructs for understanding how and why a given policy works, the potential moderator variables to consider when evaluating a given policy, and the data sources that might be useful for evaluation.

The working group proposed a common conceptual framework to guide future FCTC policy evaluation, specifying two levels of mediating variables: those specific to the policy, and those that are part of more general pathways to the outcomes of interest. It also accepted that various other factors (moderators) might affect the size of the effect, and recognized the possibility of effects incidental to those an intervention is designed to produce. Given the already well-established relationship between tobacco use and disease, and the lag time between reductions in tobacco use prevalence and observed reductions in disease outcomes, this *Handbook* (**Volume 12**) recommends that tobacco use be utilized as the appropriate endpoint for most policy evaluations. The group elaborated on the model most completely for tobacco use outcomes, but it was also applied to policies affecting product harmfulness.

Included in this *Handbook* are logic models outlining relevant constructs for evaluating the effectiveness of policies on tobacco taxation, smoke-free environments, tobacco product regulations, limits on tobacco marketing communications, product labeling, anti-tobacco public communication campaigns, and tobacco use cessation interventions. Additionally, it provides examples of measures used to assess key constructs, with special attention to measurement issues with survey methods. Also provided are descriptions of sources of data on tobacco control policies, tobacco production and trade, and repositories of youth and adult surveillance surveys. These sources of information are particularly important for making comparisons between countries, and in some cases can be used to demonstrate the impact of policies, although not the mechanisms by which they occur. Thus, **Volume 12** is offered as a guide to evaluators in the field, and consequently a frame for future IARC *Handbooks* that focus on evaluating the impacts of societal level interventions to control cancer, and other preventable diseases, caused by tobacco use.

Chapter 1
Ensuring effective evaluation of tobacco control interventions

Introduction

This volume is concerned with methods for evaluating the evidence for the effects of policy initiatives. By policies we mean the enacted decisions of governments and their consequences on the environment (legal, social and physical) in which tobacco use takes place or on tobacco use directly; that is, specific instances of the policy's manifestations (interventions). This means evaluating the effects of laws, regulations, taxes, administrative decisions, programmes and efforts to publicise or disseminate discrete interventions such as smoking cessation aids. It includes evaluation of policies that have the explicit goal of tobacco control, as well as policies that affect tobacco use incidentally, although our focus is primarily on the former. The Working Group (WG) is primarily interested in evaluating interventions that are designed to have effects at a population level, especially those enacted at a national level, but the principles apply to many subnational- and even local-level policies. While the focus of the WG is on how to assess policy consequences of governments, the evaluation framework we have developed could equally apply to the disseminated programmes of non-governmental agencies.

This chapter provides an introduction to the importance of having well-evaluated, population-level tobacco control interventions and of having a framework for achieving them. It outlines criteria used to evaluate constructs and measures, and how these relate to strategies for most effectively gathering information to evaluate the effectiveness of interventions, the mechanisms by which they work, and the conditions that moderate their effects.

Cigarette smoking is not only the most prevalent form of tobacco use, it is also among the most harmful, as it kills one in two long term users prematurely. In the 20th century, cigarette smoking caused an estimated 100 million deaths worldwide. Most of these deaths were in developed countries of the world where cigarette smoking first became popular in the 1920s to 1940s. This resulted in an epidemic of smoking-induced cancer, heart disease, and chronic obstructive pulmonary disease (COPD) deaths. In 2000, tobacco use was responsible for approximately 4.83 million deaths, evenly divided between the industrialised and non-industrialised worlds (Ezzati & Lopez, 2003). If current trends continue, it will cause some 10 million deaths each year by 2030, with around 70% in low-resource countries (Peto & Lopez, 2001; Ezzati & Lopez, 2004). This projected shift is due, in part, to increasing population size and increased smoking in low-resource countries, but it is also partly due to greater success in controlling smoking in many higher-resource countries. In the 21st century, if current usage patterns persist, smoking will cause approximately 1000 million deaths, a tenfold increase over the previous century (Gajalakshmi *et al.*, 2000). A substantial fraction of these expected deaths could be averted by efforts to discourage tobacco use and to assist those addicted to tobacco to quit (IARC, 2007a).

Most countries have ratified the World Health Organization's (WHO) Framework Convention for Tobacco Control (FCTC). It is the first piece of international law emanating from the WHO. Its objective is:

"…*to protect present and future generations from the devastating health, social, environmental and economic consequences of tobacco consumption and exposure to tobacco smoke by providing a*

framework for tobacco control measures to be implemented by the Parties at the national, regional and international levels in order to reduce continually and substantially the prevalence of tobacco use and exposure to tobacco smoke." (Article 3) (WHO, 2003).

To achieve this objective, the WHO FCTC calls for a comprehensive range of measures, specifically:

- Price and tax measures to reduce demand (Article 6)
- Protection from exposure to tobacco smoke (Article 8)
- Regulation of the contents of tobacco products (Article 9)
- Regulation of tobacco product disclosures (Article 10)
- Controls on packaging and labelling of tobacco products (Article 11)
- Programmes of education, communication, training and public awareness (Article 12)
- Bans on tobacco advertising, promotion and sponsorship (Article 13)
- Programmes to promote and assist tobacco cessation and prevent and treat tobacco dependence (Article 14)
- Elimination of illicit trade in tobacco products (Article 15)
- Measures to prevent sale of and promotion of tobacco to young people (Article 16)
- Provision of support for alternative crops to tobacco (Article 17)

In addition, Part VII of the WHO FCTC, on "Scientific and Technical Cooperation and Communication of Information" spells out a framework for research, surveillance and technical cooperation to facilitate the achievement of the policy goals.

Article 20, "Research, surveillance and exchange of information", calls for "The parties [to] undertake to develop and promote national research and to coordinate research programmes at the regional and international levels in the field of tobacco control." The article, among other things, calls for the development and promotion of national research efforts, national systems of surveillance of tobacco consumption and related social, economic and health indicators; coordination of activities so that data can be compared across countries; exchange of publicly available scientific, technical, socio-economic, commercial and legal information, as well as information regarding practices of the tobacco industry; and that the financial and institutional resources be put in place to allow this to happen.

Article 22, "Cooperation in the scientific, technical, and legal fields and provision of related expertise", expands on Article 20 with regard to such things as providing developing countries with technical and material support and training, and identifying methods for tobacco control, including comprehensive treatment for nicotine addiction.

The WHO FCTC will likely result in the proliferation of policies and associated programmes designed to reduce tobacco use. These will include but not be restricted to those mandated or recommended by the Convention. Ensuring the right mix of policies requires an understanding of the determinants of tobacco use and of how tobacco harms health.

Tobacco use is determined by multiple factors, and attempts to control the epidemic require changes in societies as well as individuals (see Figure 1.1). Analysis of the factors that influence tobacco use should encompass smokers, those vulnerable to uptake, tobacco products, those who produce and sell tobacco products, and governments who determine the parameters of use. The role of cultural and economic diversity should also be considered. Further, we need to understand how both the determinants of use and actual use and/or exposures are affected by interventions.

Policies and the disseminated programmes that result from policy decisions are of particular interest because of their potential to affect large numbers of people, in some cases entire populations. As a result, it is important to be able to show that they achieve their objectives and do so in a cost-effective way, with any incidental effects ideally having net benefits. Evaluation allows the most effective interventions to be maintained (and perhaps improved further) while less effective interventions are either improved or abandoned.

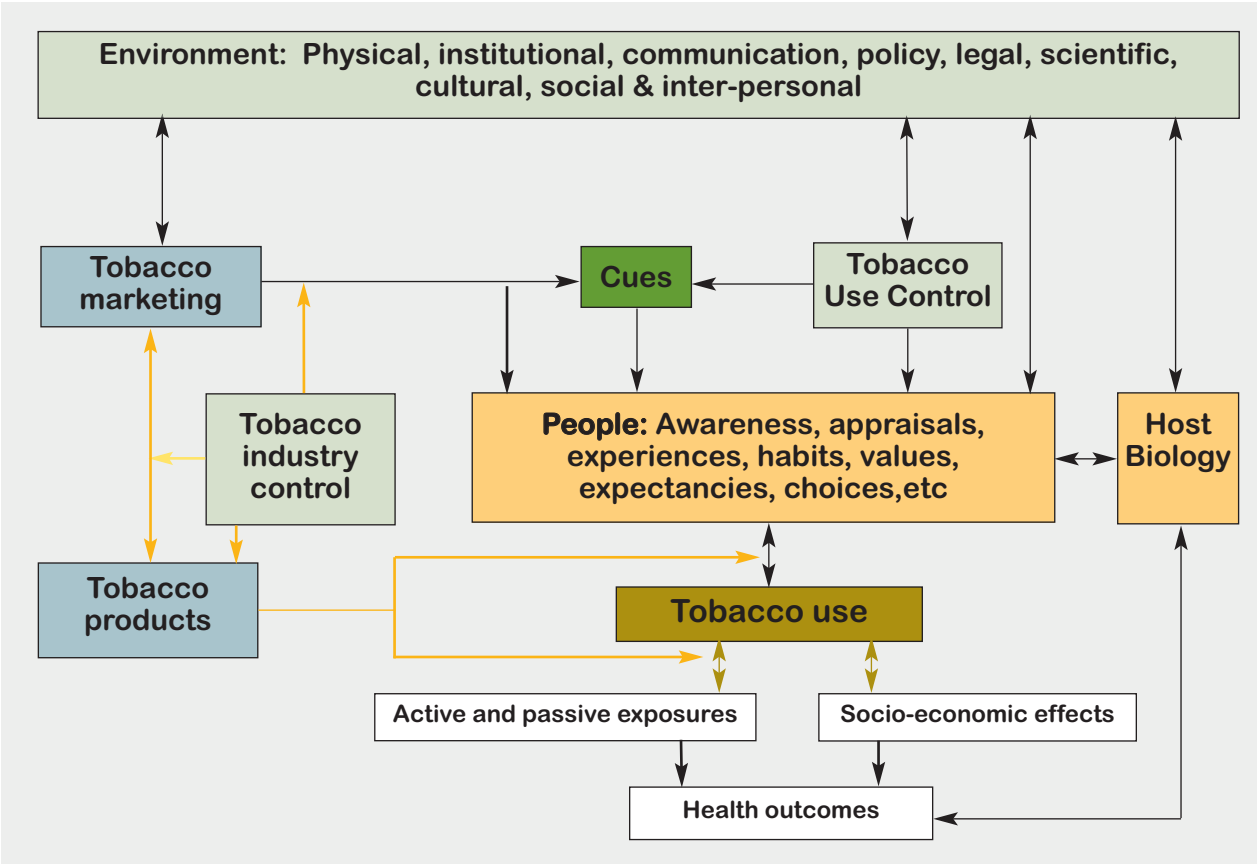


Figure 1.1 Major influences on tobacco use and its consequences
Used with permission of Ron Borland

Tobacco and health

The amount of harm created by tobacco use in a given population is a function of the toxicity of the products, the site(s) of exposure, the toxins taken in, the period over which exposures occur, and the distribution of those exposures in the population (IARC, 2004, 2007b). The harms from tobacco use are mainly from long-term use, which is made more likely by the addictive nature of the product. Calculation of the potential harms that tobacco products cause should consider the composition of what is ingested and how the products are designed to be used. Thus for combusted tobacco products, the focus needs to be on the smoke, rather than on the unburned product, although the composition of the unburned product is relevant to the extent that it influences the composition and/or density of the smoke. Mode of ingestion is often ignored; however, some chemicals are more toxic when absorbed through the lungs than through the mouth lining or stomach because the lungs are more sensitive. The evidence that exclusive cigar or pipe smokers have notably less health risk than cigarette smokers (Doll, 2004) is probably because these smokers tend to only take the smoke into their mouths.

Decades of research on the health effects of tobacco have identified numerous diseases causally related to tobacco use, including several sites of cancer (including lung, oral cavity, esophagus, larynx, stomach and pancreas), major vascular diseases

(including ischemic heart disease, peripheral vascular disease and cerebrovascular disease), major respiratory diseases (including chronic obstructive pulmonary disease, tuberculosis, and pneumonia), reproductive effects and reduced bone health. Epidemiological methods have been applied to estimate how much of these diseases in different populations with different tobacco use histories is due to tobacco (Peto *et al*., 1992).

While prolonged exposures are responsible for most fatal consequences of tobacco use, there is increasing evidence of adverse short-term effects, seen most clearly in the rapidly reversible impacts of passive smoke exposures on non-smokers (Raitakari *et al*., 1999; Wong *et al*., 1999; Wakefield *et al*., 2003a). There is no safe level of exposure to tobacco smoke. Risks of cardiovascular problems are largely reversible, and effects seem to asymptote at lower doses than those related to cancers and chronic lung conditions (e.g. emphysema), where the dose-response curve is more linear across typical exposure patterns (Law & Wald, 2003; Pechacek & Babb, 2004). The addictive nature of tobacco makes it likely that people who begin to use it will not be able to stop before the negative effects associated with long-term harm start to occur.

Nicotine is the main psychoactive ingredient of tobacco and the source of its addictiveness, but is otherwise a minor contributor to the harm (Murray *et al*., 1996; Benowitz, 1999). Most of the harm is due to other constituents in tobacco and tobacco smoke (IARC, 2004). Thus nicotine only indirectly contributes to most of the harms, by leading to prolonged use of dirty delivery systems, especially cigarettes.

The epidemiology is clear. The health risks of smoking are far greater than those associated with smokeless tobacco use. The health risk of each kind of smokeless tobacco varies significantly as a function of their toxicity. For smoked products, the likely variability in toxicity does not seem to translate into clear differences in health risks. To date, cigarettes with levels of toxins reduced by enough to be plausibly less harmful are not used by smokers, so are irrelevant to tobacco control efforts.

Some harms, particularly minor harms and those related to cardiovascular disease, are reversible on quitting smoking. While quitting can improve health, cutting down on consumption does not seem to (Hecht *et al*., 2004; Tverdal & Bjartveit, 2006). This may be in part because, for some illnesses much of the harm occurs at relatively small doses, and partly because smokers who reduce the number of cigarettes they smoke, often smoke the remaining cigarettes harder, ingesting more toxins per cigarette, thus reducing or eliminating the potential benefits of smoking less (National Cancer Institute, 2001). There has been some success in reducing the toxicity of smokeless tobacco products. Changing from smoked to smokeless products (particularly the toxin-reduced forms) can reduce harm, but does not eliminate it (Critchley & Unal, 2003; Foulds *et al*., 2003; Roth *et al.,* 2005; Henley *et al*., 2007). Reducing or eliminating smoked tobacco use is a higher priority for health than reducing smokeless tobacco use. Research is needed to determine whether smokeless tobacco might play a role in this or whether nicotine replacement products and other cessation aids are all that is needed.

Patterns of tobacco use

Tobacco is a plant containing the psychoactive and addictive drug nicotine. It has a long history of use and has been used in a wide variety of forms. The two main forms of tobacco use are by smoking and by chewing or parking wads of tobacco in the mouth and allowing the active ingredients to be absorbed (smokeless use). In the 20th century, the use of cigarettes came to dominate both the smoked and overall markets in nearly all countries. It is also the product that has been the focus of most of the research. In most countries factory-made cigarettes dominate the market; however “roll your own” cigarettes have enjoyed a resurgence in some countries. In other countries, most notably India, people consume a diverse range of tobacco products, both smoked and smokeless. Among smoked products, the “bidi” (tobacco hand-rolled in a leaf) is the predominant form used in the Indian sub-continent. Use of water pipes is common, particu-

larly in the Middle East. Cigars occupy a position as a 'luxury' tobacco product, but use is generally low. All forms of smoked tobacco are extremely dangerous to health, and there has been no major progress towards creating less toxic versions of these products that are sufficiently acceptable to consumers to be successfully marketed. Smokeless tobacco is not used in many parts of the world, but use is significant in other parts, with the products used ranging widely in places like India (e.g. gutka, use with betel quid, nicotine toothpaste), but is limited to one main form in others (e.g. snuff (powdered tobacco) either in loose or prepackaged, small tea-bag-like portions). Use of smokeless tobacco is increasing in some places (e.g. Sweden) (Foulds *et al.*, 2003). Non-cigarette tobacco use is under-researched in comparison to cigarette use.

The proportion of the population who use tobacco varies greatly from around 20% to around 60% (Shafey *et al.*, 2003). In many countries, few women smoke, often accompanied by high smoking rates in males (e.g. in Asia). By contrast, in most developed countries female smoking rates are typically only a few percentage points below that of males. There has been some predictability in these patterns of use, leading to Lopez, Collishaw & Piha's (1994) four-stage model of the tobacco epidemic, with developed countries first to experience it. In this model, female smoking initially lags behind male smoking, with female rates eventually rising. The experience of countries like Singapore and Thailand, which have so far successfully prevented female uptake, suggest that the Lopez *et al.* model does not describe a necessary progression, but that the epidemic may be able to be largely averted in some sub-populations, most notably women, when effective tobacco control policies are implemented.

Over the last 20–30 years, smoking prevalence has fallen markedly in some countries. This is well documented for some industrialised countries (Gilmore, 2000; Giovino, 2002; White *et al.*, 2003). One country, Bhutan, has banned the sale of tobacco products to its citizens. However, in some other countries, rates of tobacco use may have increased. The great diversity both between countries and within countries over time creates huge challenges and opportunities for scientific understanding. One challenge, for example, concerns preventing women from smoking in societies where few currently smoke. This challenge needs to be taken up in ways that are not contrary to the greater emancipation of women in those societies. In developed countries, e.g. in North America and Western Europe, the tobacco industry skilfully used female emancipation as a strategy for linking smoking to images of the modern woman. The slogan "You've come a long way baby" from the notoriously successful Virginia Slims advertising campaign typifies this strategy (US Department of Health and Human Services, 2001).

The most comprehensive change in tobacco control has been in attitudes and rules about smoking in enclosed public places and workplaces. As late as 20 years ago, smoking was effectively ubiquitous in most countries, with smoking allowed virtually everywhere (except where there was a danger of fires or damage to equipment). In some countries, this environment has transformed; several countries (starting with Ireland and Norway) now prohibit smoking in all public places and workplaces, and other countries are following rapidly.

The social acceptance of smoking is declining in most places where it has been studied. This decline seems to be related to the length of time the society has taken to regard the problem as serious, and to progress in the implementation of smoke-free places. In Thailand, for example, equivalent levels of smokers see their habit as non-normative (i.e., that society disapproves) as in Western countries such as Australia, Canada, the UK and the USA, all of which have decades of strong action. By contrast, even though personal disapproval of smoking is high in neighbouring Malaysia, which has only recently taken up the issue systematically, smokers are far less likely to perceive societal disapproval (ITC South East Asia project, unpublished data).

However, it is not just trends in tobacco use and tobacco-related knowledge that are likely to affect efforts to control tobacco use. Broader societal issues may also play a key role. The rapid

emergence of China and other countries as economic powerhouses is likely to affect tobacco use, at least in those countries, as more and more people have money to spend on consumer products like tobacco that are marketed to appeal to "modern" sensibilities. Worldwide concerns about the environment, including the issue of global warming, and the rise of religious fundamentalism in some countries are also likely to have effects, but it is beyond the scope of this volume to speculate as to what these effects might be. However, unless efforts are made to understand how tobacco control fits into broader social changes that are sweeping the world, important determinants of use may be missed, with the resultant reduction in the capacity to identify and implement policies and programmes that work.

In thinking about the potential health benefits of interventions, it is important to consider both their potency and their timing (see Figure 1.2). While the understanding of their potency is focal to this volume, it needs to be remembered that the sooner action is taken, the more lives will be saved. Every year of delay adds millions to the eventual burden of lives lost. Enough is known to act in a comprehensive manner now. The evaluation effort is primarily about helping us refine those interventions, to ensure they are delivered in ways that maximise their effects, and only secondarily, to the development of new more effective interventions.

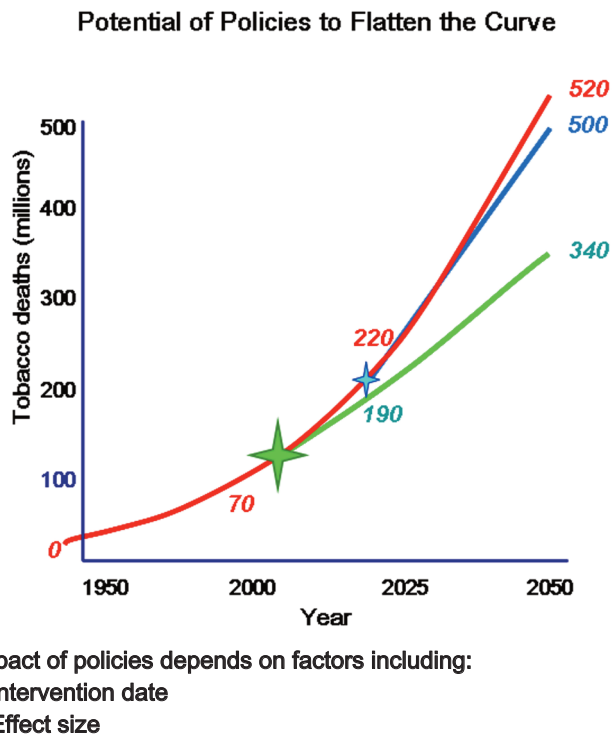


Figure 1.2 Projected impact of population-level tobacco control interventions on estimated cumulative tobacco deaths

Estimated cumulative tobacco deaths 1950-2050 showing the effects of different intervention strategies. In red baseline, in blue if proportion of young adults taking up smoking halves by 2020 and in green, if adult consumption halves by 2020

Adapted from Jha & Chaloupka (1999), The World Bank

Where does this volume fit within Tobacco Control?

This Handbook is not intended to be a one-stop resource for all tobacco control evaluation needs. It is designed to present a framework for evaluation directed at policy effects and to provide strategies and measures that are specific to tobacco control, rather than try to replicate material that is general to all forms of evaluation.

In analysing the potential contribution of research to policy evaluation, it is useful to outline the various roles it can play. Applied science proceeds through a series of iterative stages once a problem has been identified (in this case tobacco as a cause of health harm): elaboration of a theory or theories as to the cause of the problem and of possible solutions, observation and description of the problem informed by the theory, understanding causal mechanisms, intervention

development, intervention deployment and evaluation, and re-evaluation of the problem. From this, there might be the need for new or revised solutions, which may require refinement of the theory or development of a new one. Research can play a number of important roles in the process of developing and disseminating the most effective policy interventions. It can be used to:

1. help in the development of new interventions;
2. help make the case for an intervention being adopted;
3. fine-tune an intervention before implementation to meet local needs (formative evaluation);
4. document the quality of implementation (process evaluation);
5. assess the effectiveness of component parts, or of the intervention under ideal circumstances;
6. evaluate the effects of the intervention as implemented, both intended and incidental;
7. determine the cost-effectiveness of the intervention; and
8. assess the cumulative effects of changes in outcomes on health.

Of these, only number 6 is of focal interest here. All of the others are important, but to have covered them all would have made the volume too broad and too long. We also do not consider evaluation of the efficacy of discrete interventions that can readily be tested in randomised trials; e.g. smoking cessation aids. The Cochrane Collaboration (www.cochrane.org, for reviews) provides regularly updated reviews of evidence in these areas. However, we are concerned with the evaluation of effects of these interventions when applied to populations.

The focus of this volume is the evaluation of tobacco control policies in the short to medium term. We concluded that for policies directed at tobacco use, tobacco use was the outcome of interest, rather than on the subsequent health effects. Clearly, as we move forward, we will want to evaluate the summative effects of all the efforts to reduce tobacco use, and the consequential health outcomes. For a few jurisdictions that have had active tobacco control programmes for decades, this process is already underway (Thun & Jemal, 2006). However, the reality is that for most countries, we will never know exactly how many tobacco-caused deaths are being averted, because there is insufficient data on how many such deaths are currently occurring. The global estimates referred to earlier are a result of careful extrapolation from those countries where good data is available and from studies that have been able to estimate the fraction of deaths from various causes that are due to tobacco. The methods for doing this are beyond our remit, as are ways to model the potential impacts of interventions on smoking prevalence or on the burden of disease (e.g. Levy *et al.*, 2006).

The typical evaluation research study can be thought of as having five components:

1. A research design
2. The choice of constructs and measures to assess them (predictors and outcomes)
3. A sampling strategy
4. Study implementation
5. Data analysis

Of these, we only focus on the first two, although some attention is given to issues of sampling, particularly of the value of having representative samples as a core part of the research design. We do not consider data analysis as the tools here are largely generic and are covered in the main computer analysis packages, including the emerging techniques of GEE models (Generalized Estimating Equations) (Hanley *et al.*, 2003).

This Handbook was not written with the needs of those conducting evaluations at a community level in mind. However, much within it is likely to be relevant, at least at a conceptual level. The cumulative approach adopted means that for evaluations of interventions that have been shown to be effective in comparable situations, the need for intense evaluation will be less, as the evaluation can rely on indicators validated in previous work. However, for novel interventions, the more powerful methods outlined here should still be used wherever the resources allow. The US Centers for Disease Control (CDC) has published a useful guide to the evaluation of more local programmes (MacDonald *et al.*, 2001). A major difference between that guide and the present volume is the capacity to use national surveys and data collections in ways that are not

usually possible for local initiatives. That said, to evaluate local initiatives country-level data can be used as a control, with complementary data collected from the community to assess the intervention effects.

Policy areas not emphasised in this volume

There are a number of tobacco control policy domains that are either not included, or not emphasised. This is not because the WG believes that they are not important, but because it sought to keep the size of the volume manageable. Policy domains not focussed on include some that are designed to affect tobacco use directly, such as sales to minors, restrictions on sales outlets, and school-based prevention. Others are directed more at the tobacco industry, or parts of it, and include prevention of illicit trade, industry subsidisation, controls on access of for-profit companies into the market (and the role of government monopolies), and agricultural policies that affect leaf production.

The most significant area we have not focussed attention on in the volume is the lack of detailed attention to population-level prevention policies. There is a large body of evidence on the effectiveness of school-based education programmes (Thomas & Perera, 2006). The current evidence shows that, taken in isolation of other societal efforts, the impact of school-based programmes is generally weak, and there exists the potential for poorly thought-through programmes to actually be counterproductive. Most of the research on the effects of prevention programmes in schools is from industrialised countries. School programmes are plausibly of more importance in non-industrialised countries, where school is a conduit for new knowledge into the community in a way it no longer is in industrialised countries. The difficulty of developing successful prevention education comes at least in part from the problem that raising the issue engenders interest and thus curiosity about the products. Doing this in a way that overcomes the potential threat of curiosity leading to increased experimentation, and that has a net negative effect on use, has proven difficult. This may explain the interest of some tobacco companies in promoting such strategies. To the extent that educational programmes are translated into the mass media, strategies for evaluating them are covered in Section 5.6 on Measuring the Impact of Anti-Tobacco Public Communication Campaigns.

Another prevention strategy we do not address the evaluation of is policies to prohibit sales of tobacco products to minors, and to enforce these laws by using young people attempting purchases. Such programmes can result in a decline in the proportion of such attempts that result in sales, but the evidence that this actually reduces youth smoking is not strong (Stead & Lancaster, 2000).

In the broad area of tobacco industry control, there is some consideration of illicit trade in the section on sources of production and trade (Section 4.2) and in the section on tax policies (Section 5.1). Neglected areas include restrictions on the number or type of outlets in which products are sold. There are few examples of attempts to restrict the number or type of outlet selling tobacco. However, it seems inevitable that in the future some jurisdictions will try to restrict access to all smokers, not just youth.

We also do not address the evaluation of policies that restrict for-profit companies from operating in the market. Some countries have actual or virtual state monopolies on the sale or production of tobacco products. Several countries have been forced to abandon these monopolies by the World Trade Organisation. It has been argued that non-profit control of the industry should make tobacco control efforts easier (Borland, 2003), but there is little work evaluating either the move to free markets or the potential of restricting the markets. In both these areas, research is needed to evaluate possible options and to estimate likely effects.

A critique of current approaches to evaluation

To achieve maximally effective tobacco control requires the development and ongoing refinement of a viable set of

methods for integrating research and evaluation in the implementation of tobacco control interventions. The population health challenge is to use scientific methods to ensure that systems are set up to understand the effects of the policy initiatives in such a way as to allow their evolution into the most effective ways of controlling the epidemic of tobacco use and related harms. Evaluation researchers in tobacco control, like professionals in other areas of population health, have been concerned for some time about limitations in the evaluation framework used.

The current dominant model of intervention evaluation for improving population health involves extrapolation from the use of randomised controlled trials (RCTs) of clinical (most typically, pharmaceutical) therapies. It is based on the desire to identify the active therapeutic agent or agents within any intervention. This model is important and extremely successful for testing the efficacy and often effectiveness of discrete interventions offered at the individual (and even small group) levels, particularly where double blinding is possible. This is where neither researcher nor participant know who is getting the therapeutic agent under evaluation and who is getting either a placebo or the existing best-practice intervention. RCTs produce considerable certainty about causes. However, reliance on RCTs is not always possible or appropriate for the evaluation of policy impact in the population for a number of related reasons. First, implemented policies cannot be randomised and analogue studies, where randomisation can occur, may lack critical elements of policy interventions (e.g. authority of law, or it being applied to all in the community). Second, over-reliance on RCTs, which focus on the detection of intervention effects, can lead to a neglect of theory, which is critical for generalising from results to related areas, and for understanding the mechanisms by which interventions work. Third, RCTs are not able to answer questions about the relative effectiveness of interventions across different populations. Fourthly, when RCTs are compromised, in terms of deviation from the double-blinded ideal, they are less powerful, and may be less strong than alternative methods with different validity limitations. Finally, focussing on RCTs to provide answers to questions can result in a neglect of other evaluation techniques, which although not as inferentially strong as RCTs, may have complementary strengths. It is important to understand the conditions under which RCTs are limited and what the implications are for inference.

Limitations of RCTs

Determining whether a discrete intervention works involves answering three questions, which sometimes can only be answered separately: the questions of efficacy, effectiveness, and dissemination (Flay *et al.*, 2005). First is the efficacy question: Can this intervention work? That is, when implemented in a controlled and optimal way, does it work? Here the double-blinded randomised controlled trial (RCT) is the gold standard, where possible. Second is the question of effectiveness: does it actually work when implemented under real-world conditions, and with what degree of variation? Third is the question of dissemination: Is the intervention used by enough of the population who would benefit from it to have an impact? An effective intervention that few are prepared to offer or few are prepared to use is of little benefit. One must also consider the extent to which the intervention is similarly attractive for all with the problem. When only a subset of the population benefits, any barriers to selective adoption or influence should be examined. As we move from addressing questions of efficacy, through effectiveness, to dissemination issues, it becomes increasingly difficult to fit the conditions for RCTs, even for clinical interventions.

RCTs involve a number of (usually implicit) assumptions. First, RCTs assume that the measurement required for the evaluation does not affect the integrity of the intervention. Second, it is presumed that the interventions can be evaluated in isolation of environmental factors, including the society's response to

the intervention and to other cultural trends; i.e., that the effectiveness of the intervention can be determined prior to its widespread implementation. Third, it is assumed that any impact of personal choice over whether to have the intervention can be separated from the core therapeutic effect. Fourth, it is assumed that the intervention is uniformly effective for all who are eligible to be given it. None of these assumptions are tenable for policy interventions and disseminated programmes.

The assumption that a given dose of an intervention is assumed to have an equivalent effect on all who have the condition it is intended to treat is problematic even with many pharmaceuticals. The solution to this problem has been to treat each identified population as novel and to require separate RCTs. This might work for major distinct differences, but when there are many possible populations to consider, the strategy becomes cumbersome and costly. More efficient strategies are required.

RCTs are similarly a cumbersome method for evaluating interventions that vary continuously, as they involve creating discrete categories for randomisation. This means there is, for example, poor quality information on optimal dosage, both amount per dose and duration of use. This makes RCTs a particularly cumbersome method for evaluating interventions where the dose of an intervention can vary considerably.

Finally, there is no capacity to consider closely related — indeed, functionally equivalent — interventions as a class, and develop different criteria for evaluating new versions of essentially the same intervention. For example, different executions of a cognitive-behavioural cessation treatment or even the various forms of Nicotine Replacement Therapy (NRT) get treated as independent interventions. In the case of NRT, all variants have had to go through the same process of testing through independent randomised trials before they were able to be marketed.

Population interventions tend to be different in observable ways wherever they are implemented. Information-based interventions are dependent on language, and the language used must vary by culture, not just linguistic group. Language must be kept up-to-date to make it contemporary, and thus attract interest (and sometimes increase) comprehension. People-based interventions invariably differ. Policy-related interventions encompass those major aspects of the system that allow them to operate, not just the core requirements. It is not reasonable to assume that population-based interventions have their effects independent of anything the person does or thinks, unlike most pharmaceutical interventions. Like virtually all psychological and social interventions, as well as some pharmaceutical and other ones, the effectiveness of policy interventions is critically dependent on how the individual responds to them. For clinical interventions, the frame is quite different. Their questions are framed: If the appropriate system is put in place to ensure the person with the illness uses the intervention properly (or is given it properly), then can we demonstrate a benefit? The question the WG ask is quite different and much broader: Can a system be put in place that will make the intervention work, and how can that system be optimised under different conditions?

Where limitations exist on the internal validity of RCTs for making the inferences of interest, the strategy of using meta-analyses of similar studies to draw inferences is similarly problematic. Alternative means are required to control for these threats to inference. It is only in the context of being able to assume generality, having few enough interventions to assume each is an independent case, and having the capacity to test interventions in isolation of their context, that the model of RCTs as the keystone of evaluation is possible.

The allure of having a simple model based on RCTs to allow definitive inferences about the effects of interventions treated in isolation seems to have distracted us from considering the potential utility of other approaches. In particular, the RCT-focussed framework tends to neglect the role of theory and of the potential contribution of combined studies with different kinds of limitations.

The contribution of theory is undervalued in tobacco control and in public health more generally. We agree with the noted psychologist Kurt Lewin: "There is nothing so practical as a good theory." Some in the social sciences take theory to refer to the existing, sometimes demonstrably limited social science models, and take the theories from other areas (typically from the biological sciences) to be accepted fact, rather than theoretical models; e.g. of how a chemical will affect behaviour. Theory is thought of in an encompassing sense of the accumulation of our understanding of how things work, not merely the original ideas. Theory provides the mechanism to systematically use existing knowledge to understand likely future effects. The aim should be to develop consistent sets of ideas (theories) that describe and predict actual outcomes. A hunch or a past empirical finding is an unarticulated theory of what will happen in the future. Unless articulated, these implicit theories cannot be subject to proper scrutiny. If they turn out to predict outcomes, there is no capacity to work out why without first articulating them.

Theories specify mechanisms or mediating pathways of effects, allowing these pathways to be tested. They also can specify conditions under which interventions will work (i.e. moderate intervention impact). One can test whether these factors affect outcomes, and thus be better placed to develop the suite of interventions needed to provide maximal help to all, or to produce the desired structural or cultural changes. No single theory can encompass the complexity of controlling tobacco use; however, more can be done to consider how theories that deal with different aspects of the problem interrelate, including different timescales for change (e.g. behaviour change versus change in cultural norms and practices). The set of theories used should be compatible with each other, even if the nature of the interrelationships is not fully articulated.

The most important implication of considering theory is that it allows explicit linkage of tobacco control to relevant existing knowledge. A focus on evaluating interventions in isolation tends to distract from what is known, specifically:

- Information campaigns can increase knowledge about tobacco.
- Knowledge can change beliefs and attitudes.
- Beliefs and attitudes can affect tobacco use.
- Advertising can change behaviour independent of conscious awareness of the influence.
- There are programmes and aids that can help people quit using tobacco.
- There are ways that the toxicity of products can be reduced.
- Price rises affect levels of consumption of tobacco products.
- Poorly designed and/or executed communications can have boomerang effects.

This knowledge is part of a foundation that is sometimes forgotten. The question we are really asking is: Under what conditions can the desired effects be optimised? This includes concern about the form of the intervention, the ways it is delivered, and various characteristics of the populations to whom it is provided.

A new evaluation framework, one that is less reliant on the RCT, is required. It should have a systems perspective; use the best possible methods, including RCTs where appropriate; allow a more central role for theory, to allow more efficient consideration of possible variation in effects across populations; and provide a more efficient means of understanding effects of dosing and other aspects of implementation.

One approach to evaluation that is popular among public health practitioners, but that has less credibility with researchers, is that of programme evaluation (e.g. Patton, 1997). These models have grown in areas where there are no simple relationships between programmes and sought policy outcomes, yet there is a need to demonstrate progress. Thus the focus of these models of program evaluation is often on determining intermediate effects when it is difficult to demonstrate effects on the main outcome goals. We believe that there is value in extending these models to consideration of outcomes as well. The essence of these approaches is to test the theory behind the programme, sometimes also

called the "programme logic", to assess whether the various aspects of a programme can be shown to contribute to the achievement of its goals (MacDonald *et al.*, 2001). The WG has adopted the idea of using logic models as a core element of the framework we have developed. We found that doing so increased conceptual clarity and provided a useful organising frame for thinking about the policies and a more coherent way to organise the chapters and sections.

Framework for tobacco control evaluation

The role of evaluation is to determine the effects of interventions, determine under what circumstances these effects occur, and help identify ways to make the interventions more effective. To do this involves determining how the interventions work, and diagnosing any problems that either prevent them from working as desired or diminish their impact, in particular any differences of effects within the target population (equity issues). In doing this one should consider the totality of effects, both intended and incidental. To do effective evaluation we need to consider what effects might occur (theory), and design studies that allow detection of effects in the variables of interest (description) and making of valid causal inferences about the contribution of the intervention to the observed changes in outcomes.

Theory

Evaluation must begin with a theoretical evaluation of how an intervention might work. Often there will be one clear theoretical mechanism, generally provided as part of the justification of having the policy, but sometimes alternative modes of effect might be postulated. This is particularly the case when the head of power (constitutional source of capacity to legislate/regulate) under which policies are enacted is limited. Thus policies to protect workers from exposure to passive smoking cannot explicitly consider the possible benefits of smoke-free places for reducing cigarette consumption or for enhancing quitting. Good evaluation requires consideration of all potentially important outcomes, not just those used to justify or provide a legal basis for the policy.

Evaluation is enhanced by showing the mechanisms of the effects, not just restricting itself to determination of effect size. This is critical in population research because most of the outcomes we are interested in are potentially determined by multiple factors; thus it helps demonstrate a contribution from the focal interventions as distinct from other interventions happening at the same time. Thus, the theory needs to spell out the mediational model of how an intervention might work. Mediational models allow us to test each step along a proposed causal chain from intervention exposure to behaviour (see Figure 1.3). If some relationships are not as predicted, the intervention may not be working, at least in the way it was intended to work. In cases where the intervention is known to be potent, evaluation of mediators may only need to proceed as far as assessing uptake/exposure. However, where the potency is unproven, testing the intervention's impact through to the desired outcomes (e.g. smoking cessation) becomes necessary. In an area like tobacco control where the main outcomes of interest (e.g. smoking cessation, pre-

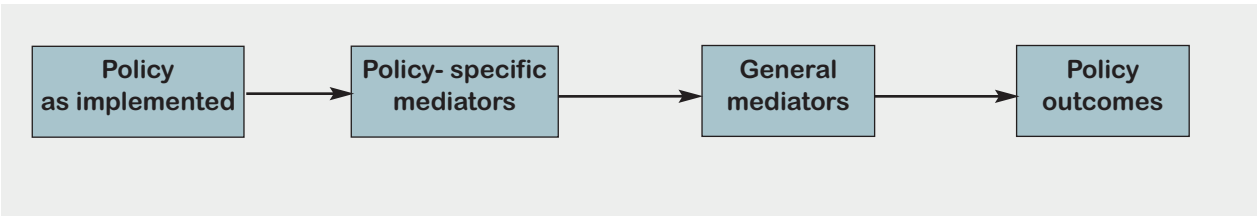


Figure 1.3 A generalised model of mediation

vention of uptake) are determined by multiple factors, mediational models can also help establish the relative contribution of specific interventions. Testing mediational models can also enhance understanding of basic mechanisms and facilitate the development of new and improved interventions.

Other theoretically important factors are those that may moderate the relationship between the intervention and outcomes. That is, what conditions affect the efficacy of the intervention, or how does its effectiveness vary by identifiable sub-groups. Where one finds or theorizes moderator effects, it is important to understand where they occur along the proposed mediational pathways, or indeed whether different mediational pathways exist for different groups or situations (see Figure 1.4). For example, if an intervention is not seen to be relevant to or targeted at a group, this group may not respond to it. Here, making the intervention relevant might be all that is needed to remove the moderating effect. A good example of this is advertisements whose spokespeople are old, which are typically not seen as relevant to young people (the converse is less likely to be true). Something as simple as choice of actor can create moderator effects, which under other conditions would not be present (or be so small as to be ignored).

Incidental effects must also be considered. Sometimes it can be useful to separate these out from the intended effects (see Figure 1.5). Incidental effects can occur for a range of reasons; some may be theoretically expected, while others may not. Some can occur as a result of counter-actions of sections of the tobacco industry to reduce the threats of policies to their profitability. These effects can be incorporated within the more general model (Figure 1.4) as all such effects can be either due to reactions to the policy, or to independent other factors (and thus should be treated as moderators).

Description

The relevant theory tells us which constructs to measure. Evaluation requires a good description of the problem and its context, and of how these are changing. This involves finding appropriate measures of the constructs of interest and of collecting data using the appropriate measures. The goal here is to provide population estimates of what people do and think, focusing on key outcomes. It involves collecting data in four principal domains: 1) who uses tobacco, what they use, how much, and where and when they use it, as well as any relevant knowledge, beliefs and attitudes (including those of ex-smokers and non-smokers); 2) tobacco industry behaviour, including characteristics of their products; 3) tobacco control activities to which people are exposed; and 4) aspects of the broader environment that might affect tobacco use or tobacco harm outcomes (cultural norms, controls on activities like alcohol consumption that are linked to tobacco use). High-quality data collections, such as regular cross-sectional surveys, are essential to describing the nature of the problem and the

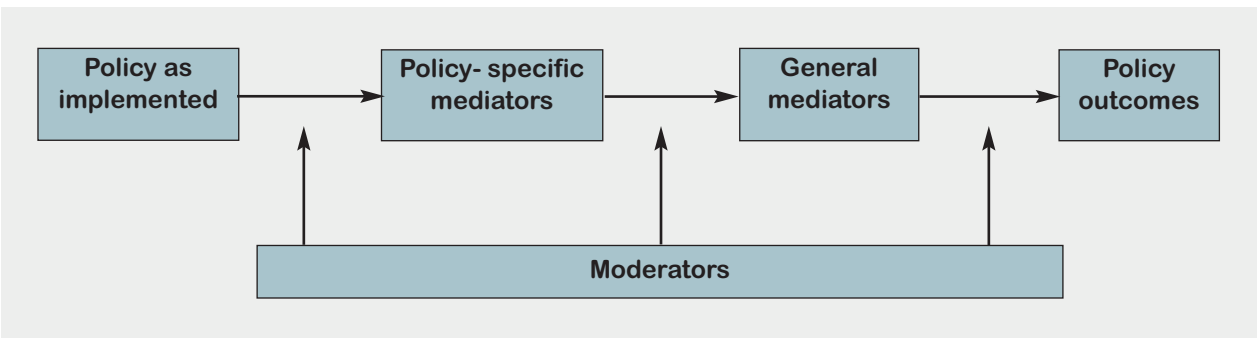


Figure 1.4 A generalised model of mediation, making allowance for moderator effects

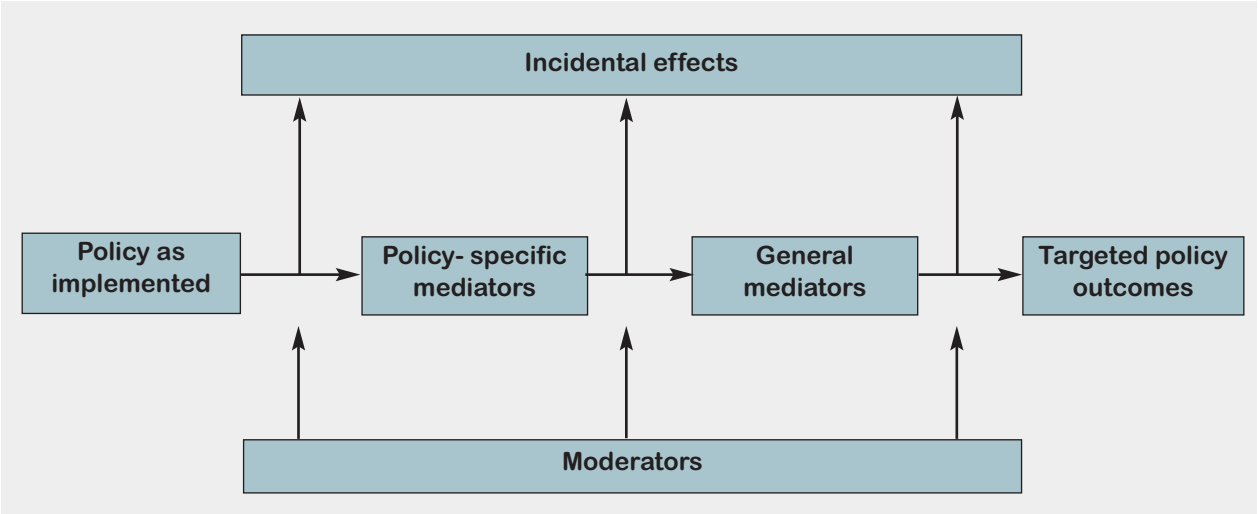


Figure 1.5 A generalised model of mediation, making allowance for both moderator and unintended or incidental effects

quantification of trends in tobacco use and in key determinants of use. In tobacco control, because the tobacco industry or sections of it might be motivated to moderate the effects of policies, it is important to conduct surveillance of possible counteractions to policies. More generally, possible incidental effects of policies should always be considered and measured where appropriate.

There are five broad types of outcomes that relate to individuals: improvements in knowledge, changes to attitudes and related normative beliefs, changes to behaviour patterns, changes in exposures, and health outcomes (particularly acute ones that can be detected soon after a policy is implemented). Interventions typically change the environmental conditions that affect and thus sustain these outcomes. Mechanisms for behaviour change can be through rules and restrictions, making available alternatives or substitutes, and/or providing relevant resources and/or skills. The mediational pathways vary both for outcomes and policies. For example, mediational pathways to knowledge acquisition are shorter than ones to smoking cessation.

Inference

The core of good evaluation is designing studies to detect changes in outcomes that might be attributable to a specific intervention, and putting in place measures to rule out alternative explanations. These alternative explanations are of three types: those related to systematic errors of measurement (bias), those related to alternative mechanisms of effect (confounding), and chance effects. Bias occurs where the measures used to assess the constructs of interest actually measure something different (usually a closely related construct) or are contaminated by some systematic error (e.g. social desirability can affect responses about beliefs and intentions). Confounding occurs when the association with the outcome of interest appears stronger or weaker than it truly is as a result of an uncontrolled association between the intervention and other mechanisms of effect (e.g. a different policy intervention). The contribution of chance is a function of naturally occurring variability in outcomes of interest, and its impact is controlled for by ensuring adequate sample sizes.

The quality of evidence from any single study is a joint function of the study design and of the quality of the measures used: that is, their reliability and validity. Where optimal research designs

are not available, one must focus on the relative strengths of different designs. It is not enough to conduct meta-analyses of the individually strongest studies. A diversity of research designs (and associated measures) with complementary strengths, should be combined, and that information combined in ways that increase the validity of inferences. Demonstration of similar effects with different methods and/or measures increases confidence in the reality of effects and of the plausible causal mechanisms.

Evaluation as a dynamic process

The evaluation of policy interventions occurs after they are instituted, as they first must be implemented somewhere before it is known how they actually work. Because the authority of government policy or law may affect compliance, it is not possible to confidently generalise from the results of analogue studies conducted prior to implementation. This means one cannot in principle be certain of the effectiveness of interventions before they are implemented; hence, lack of evidence needs to be used with caution as a reason for delaying needed policy change. Scientific methods can be used to help us minimise our risk of error, but they can never eliminate it completely. Science should not inhibit action when there is a need for action, but rather act to maximise the chances of success and minimise the risks of wasting resources. This involves a model in which science plays a role of evaluating interventions once they are disseminated, not just restricting its activity to evaluating interventions before they are disseminated. It is a science of evidence in action, not just of evidence preceding action. One aim of this volume is to provide the conceptual framework and some of the tools to allow more effective evaluation of implemented policies and programmes. It is designed to complement the often (necessarily) limited evaluation of interventions that occurs before they are implemented.

There is the possibility that empirical work will show the theoretical model used to develop and or evaluate the intervention to be flawed: either incorrect in some of its assertions (including inclusion of factors that have little or no influence), or incomplete by ignoring important factors. It is only by specifying models that one can systematically work to make them better.

A model of evaluation is required that is designed for the dynamic, ever-changing world in which we live. The potential of the world's diversity must be viewed as a tool to aid in understanding, not an obstacle to be overcome. Each action of government is an attempt to influence outcomes in ways consistent with policy goals, which, hopefully, aim to improve the health and well-being of the community. Similarly, the actions of tobacco companies are also designed to affect smoking, in this case in ways that enhance shareholder value, which is why they are almost invariably directed at increasing or at least maintaining use. Even the best thought-through interventions sometimes fail to work as expected, and policies that work in one context sometimes stop working when the context changes. Because neither past experience nor theory can be relied upon to always deliver the best solution to our problems, methods must be established to check when and how things are working. This is what modern evaluation is about. A framework for effectively evaluating policy interventions is essential.

Such a model places less stringent tests on demonstrating that something has equivalent effects in a new context or when delivered in a new form (where there is no reason to expect changes in efficacy) than it does for evaluation of truly novel interventions or their implementation under conditions where differences in effects is plausible. However, it still calls for stronger evaluation methods when evidence accumulated to question an assumption of equivalence. Thus it provides an explicit link between the roles of ongoing auditing of programmes to ensure continued effectiveness and more intensive evaluation activity when there are concerns. As these decisions are based around clearly articulated theories, the framework is open to scrutiny and should allow the most cost-effective possible evaluation by demanding plausible reasons

before testing for differences in effects.

Characteristics of interventions

Typically, policy interventions are designed to have sustained effects, but in some cases this may require designing ongoing programmes to ensure that this happens. Further, there may be short-term onset effects. For example, there is evidence that warning labels on cigarette packs have an onset effect as well as a sustained effect (Hammond *et al.,* 2007a). We need evaluation methods that can differentiate onset effects from sustained effects, and which also can help us understand the conditions under which both kinds of effects are maximised.

It is necessary to understand what, if anything, is required to sustain potential enduring effects: that is, what endures without further intervention and what requires regular updating, or a sustained presence. For example, anti-smoking mass media campaigns have a short-term impact on quitting (Snyder, 2001). It seems important to maintain cues in the environment to remind people of information for that information to have a maximal impact. The form of some kinds of interventions may also need to change over time if the effects of the intervention are to be sustained. This applies particularly to communication-based interventions. What is seen as up-to-date, and thus of most relevance for communication, changes quite rapidly in some communities. Similarly, across cultures, intervention may need to be framed differently to ensure cultural relevance. Under some circumstances it can be useful to conceptually separate the core concepts in an intervention from the mode of communication used to convey them. Thus evaluation might focus on the cultural relevance of the intervention or on its underlying potency, or both. Analogous to the way societies and/or people change, interventions need to change to maintain their relevance. This calls for an equivalent model to that of how to create new immunizations for emerging strains of influenza. Here, the rate of change in the problem is too rapid for RCTs to be practical, and quite different methods are used.

Changes to interventions may also be required as a society progresses through the adoption of an innovation cycle for adopting new sets of values and behavioural options for tobacco use. Take, for example, encouraging the adoption of smoke-free homes. This happens first in the face of social disapproval, or at least lack of understanding. An entity instituting a ban will often be asked to justify it, and some might see it as unreasonable. However, as such bans become more common, there comes a tipping point where smoke-free environments become the norm. Since justification is no longer necessary, smokers often just do not smoke when indoors, and those without such bans feel a need to justify their positions. Before the tipping point, even quite intense interventions may have limited impact (as has been the case for implementing smoke-free homes (Hovell *et al.*, 2000)), while after it people may be readily able to change without help (as evidenced by rapid adoption of the practice in some countries (e.g. Borland *et al.*, 1999)). Where things change, the rate of change must be considered as well. When it is more rapid than the time for the institutionalisation of interventions through traditional testing of efficacy and so on, then new methods must be adopted to allow interventions to be changed in train with the changing context. This is one of the reasons why it is important to pre-test the messages used for cultural relevance, even for proven interventions when applied in new contexts. This is also why it is important to conduct ongoing evaluation of disseminated interventions.

How policy interventions that target behaviour work

Evaluations of population-level interventions are typically interested in determining the overall effect of the intervention. As a consequence, it is not so much about asking whether an intervention of this kind can work, but of asking under what circumstances does it work and how to optimise those conditions to get maximal impact. This involves consideration of the reach of the intervention (sometimes no more than awareness),

the ways people respond to it and its underlying potency or efficacy.

There are three key aspects of interventions from the perspective of the individual: awareness of, acceptance of, and actions taken in response to policies. Evaluation must deal with all three. The first aspect is determining the extent to which the target population is aware of the intervention, which is a function of its implementation, dissemination, and surrounding publicity about it. Awareness is generally a prerequisite of policy effects, except in those rare cases where the policy creates environmental conditions that can have direct conditioned effects; i.e. independent of conscious awareness.

The second aspect is documenting attitudes towards the intervention by the target population, as this can affect their responses to it. Policies that are unpopular are more likely to be resisted, and forms of assistance that seen as inappropriate to the person's needs are unlikely to be adopted. Thus, a smoker who objects to smoke-free rules is more likely to ignore the rules or to seek convenient alternatives, while a smoker who approves and sees this as an opportunity to gain greater control over their smoking, may not only comply, but use the opportunity to either quit altogether or reduce their consumption. A price increase will only cause smokers to try to quit if they see the increased price as making smoking no longer worth the cost. Alternatively they could smoke more of each cigarette to maintain the value, or shift to a cheaper brand, or seek out sources of cheaper cigarettes, or even re-interpret smoking as something more exclusive and thus desirable. Like awareness, acceptance can only really be evaluated at a population level, although it is typically the acceptance of each individual that is critical. In some collectivist cultures, the views of community leaders are also critical, as they determine what it is acceptable to think and do. These roles are in addition to the roles of leaders in all cultures as policy makers.

The third aspect is the evaluation of the actions that result: that is, the consequences or outcomes of the intervention in terms of both intended and unintended incidental effects. This is a function of both the actions taken by the individual and the potency of the intervention. While traditional intervention evaluation restricts its focus on outcomes among those who are encouraged to use the interventions, for policy interventions this is not a useful restriction; one must consider the total impact on the population, including those who are unaffected. Outcomes should be considered as a joint function of the potency of the interventions, the ways they are used or responded to (a function of attitudes to them), and the degree of exposure to them.

The theories behind tobacco control

A critical step in developing an evaluation framework is having a coherent theory or set of theories as to what tobacco control is about. This should extend beyond the list of tasks identified in the WHO FCTC to an analysis of how the various domains of intervention are theorised to contribute to the overall goal. The nature of the relationship between tobacco use and harm must be sufficiently understood to know what behavioural aims are appropriate. Such an analysis should consider the broad scope of potential impacts, not just those that are part of the rationale for implementing any particular policy initiative. For example, the impact of smoke-free places, introduced to protect non-smokers, also have beneficial effects on smokers and do not appear to have some of the adverse effects on economic activity that some had feared (Scollo *et al.*, 2003). Detailed analysis of the conceptual foundations of specific interventions is provided in the relevant sections later in this volume. Here the WG addresses a few broader issues.

A broad schematic overview of key influences on tobacco use and tobacco-related harm is provided in Figure 1.1. This figure makes it clear that policy and socio-cultural influences have indirect effects on use and that the most proximal determinants of use are the product; cues in the environment; characteristics of people, including cognitions about the products; and the person's biology (both conditioned and innate). Further, the behaviour and the product jointly determine exposures, which, in interaction with existing biology,

determine harm (see Figure 1.6). The role of a systematic science of tobacco control is to analyse and clarify the components of this system and their interrelationships over time, with the aim of introducing interventions that will minimise the harms. Figure 1.6 is a generic model for this. It is possible to elaborate this figure to include other impacts of policies (see Figure 1.7). With generic models of this kind, areas that require greater attention can be expanded upon and boxes where things are more straightforward can be combined.

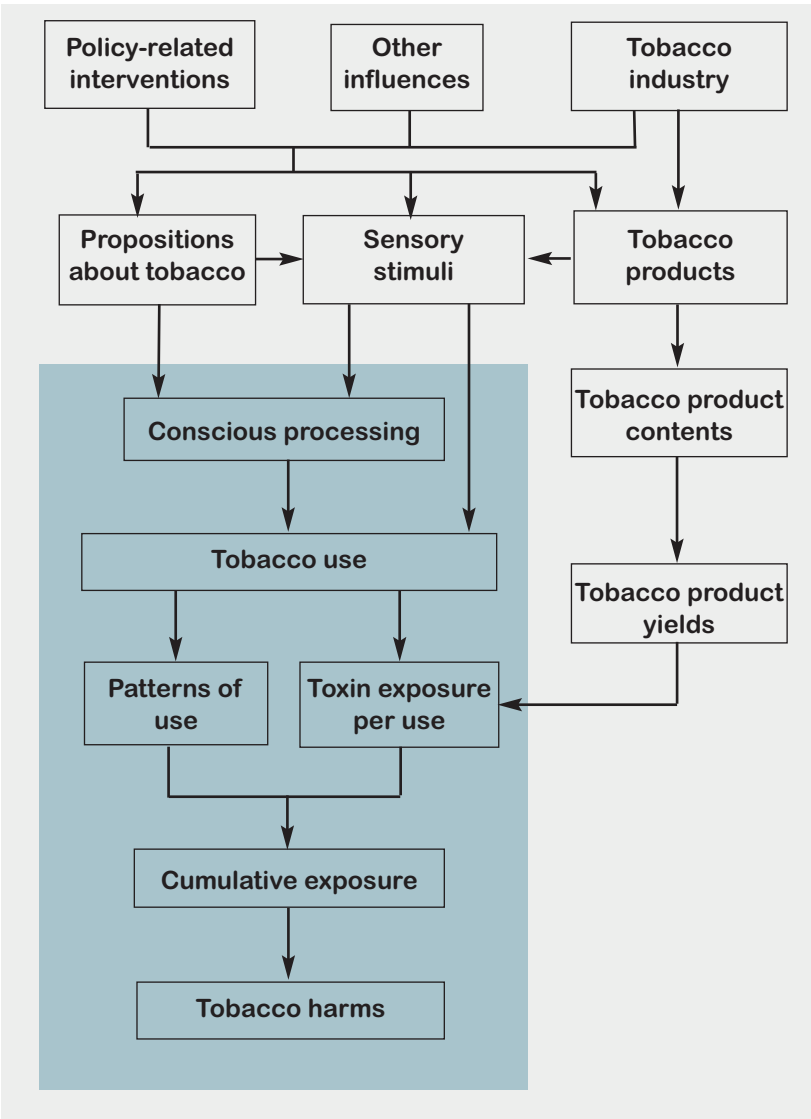


Figure 1.6 Schematic diagram of main pathways by which policies affect tobacco use, tobacco exposures and tobacco harms

Tobacco control efforts can be focussed on users and potential users of tobacco products (e.g. changing knowledge and beliefs), or they can be designed to directly reduce use (e.g. price and availability controls), or to reduce use indirectly by changing the environment to increase cues to inhibit use (e.g. warning labels on packs), or reduce cues to use (e.g. by constraining tobacco companies' marketing practices), or by changing the nature of the tobacco products on the market (see Figure 1.8). Efforts can also be directed at reducing the toxicity of tobacco products (targeting the industry), and at reducing the exposures of non-smokers (targeting tobacco users). To intervene in any of these ways with either people or companies requires a good understanding (theory) of how the factors producing unwanted effects operate and how the intervention will affect those operations. It is beyond the scope of this volume to spell out such a complex theory, although in each section, relevant elements are canvassed.

Tobacco industry controls

Tobacco industry controls are about targeting the 4 Ps of marketing: Product, Price, Place (or availability) and Promotion; to which a fifth P can be added, Packaging; and, unrelated to marketing, the imposition of specific obligations to provide information (for example, warning material) regardless of its impact on the marketability of the products. This is

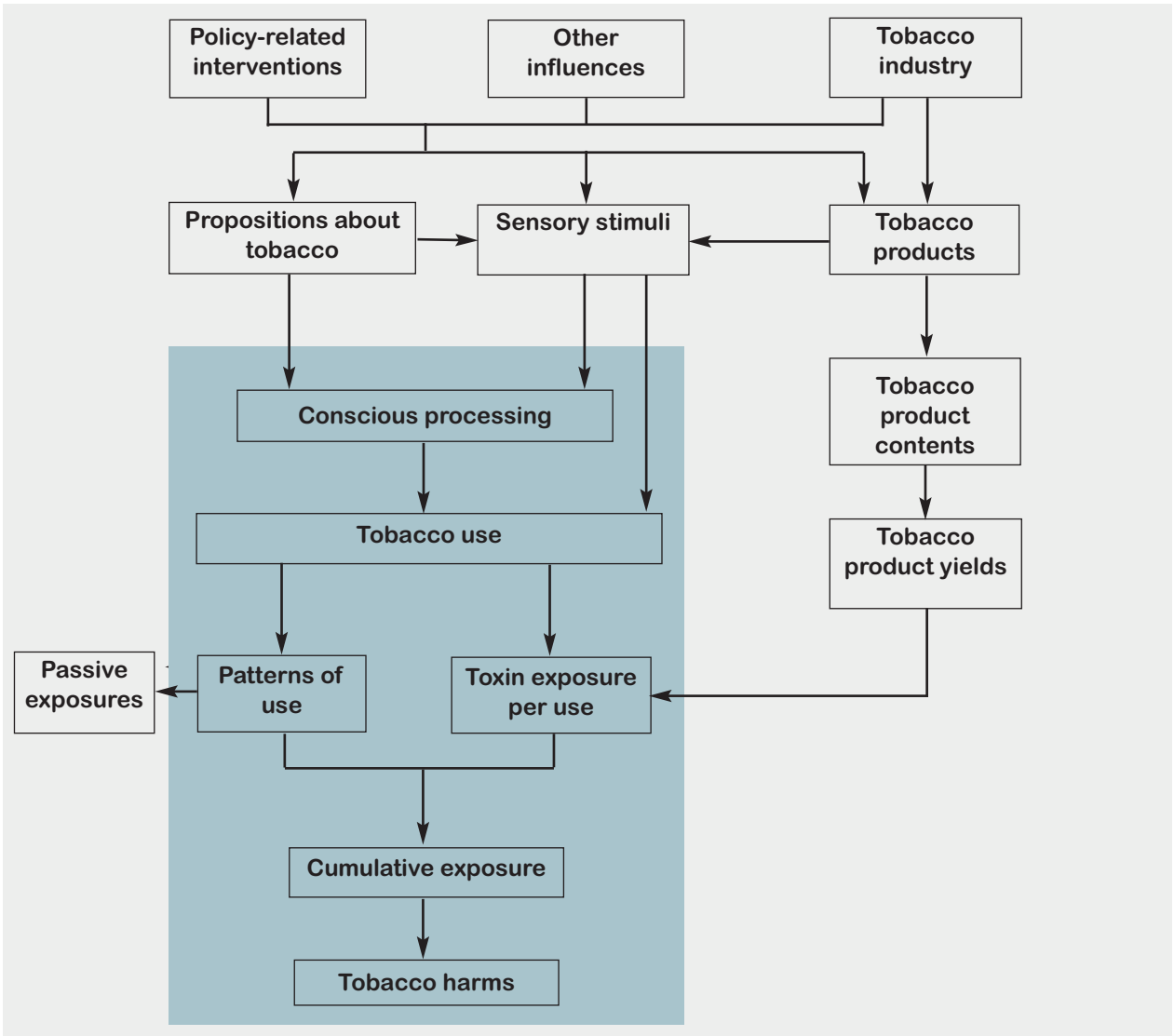


Figure 1.7 Model from Figure 1.6 expanded to illustrate where effects other than on tobacco use fit in

achieved through a mix of laws and agreements, generally targeted at manufacturers or distributors, but in other cases, at other points in the supply chain (e.g. retailers). Evaluation of tobacco industry controls also requires an analysis of possible industry actions to counter the intended effects, or to otherwise minimise adverse effects on their business.

Product controls (see Section 5.3) include rules about what types of products can be sold (e.g. smokeless tobacco is banned from sale in some jurisdiction), levels of constituents or emissions (e.g. upper limits on tar, nicotine and carbon monoxide as measured by ISO standard testing; restrictions on additives/ ingredients), or on engineering features (e.g. mandating reduced ignition propensity cigarettes, filters). The aims of

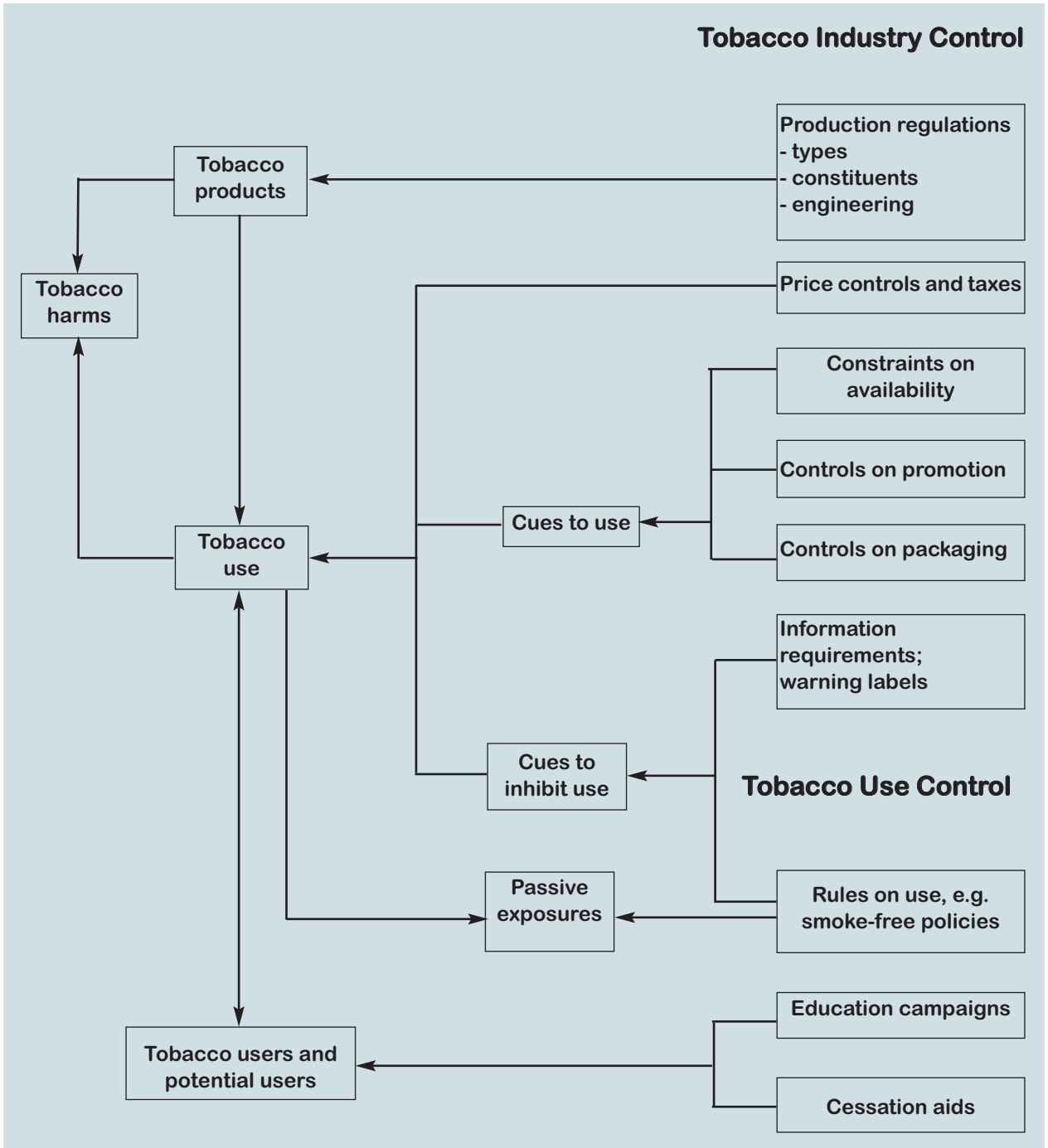


Figure 1.8 Schematic overview of tobacco control interventions and how they relate to tobacco products, users and potential users

product rules vary from preventing new forms of tobacco (to a market) becoming established (e.g. bans on smokeless), to reducing their appeal (e.g. bans on flavourings), both of which are designed to reduce use, and rules to reduce the harmfulness of the products (e.g. constituent limits), which can also have direct effects on the harm caused.

Price controls (see Section 5.1) includes efforts to damper demand through increasing prices (e.g. taxation of various forms), which can have direct effects on use, as well as strategies to prevent price-related marketing (e.g. setting minimum and/or maximum prices to prevent discounting and other forms of price-related marketing).

Place or availability controls refer to efforts to reduce the availability of the products and include restrictions on the number or types of outlets, and to whom they can be sold (e.g. age limits and bans on vending machines). Many of the existing rules have been put in place to discourage use by young people, but restrictions could also be used to reduce impulsive purchases and/or to discourage use in certain venues (e.g. bans on sales in bars).

Packaging controls include rules about what can be on the pack (e.g. use of terms like "Light" and "Mild"; see Section 5.5). It also includes rules that prohibit sale of single cigarettes and establish a minimum pack size to stop use of packs with small numbers of cigarettes, which are known to appeal primarily to young people (Wilson *et al.*, 1987; Assunta & Chapman, 2004a; Prokhorov *et al.*, 2006). The effects of such policies may operate through reducing cues to use, or by making the product less attractive, reduce the value of using such products.

Controls on promotion (see Section 5.4) are the most prominent form of control on the industry. They are essentially about reducing cues to use, but in doing so, might also reduce the appeal of the products. Controls include bans on paid advertising, sponsorships, and product placement, and encompass restrictions on packaging (including controls on the use of trademarks, e.g. generic packaging). Because tobacco is sold in a competitive market, some signs differentiating products as belonging to a manufacturer/marketer are necessary. Even in places when brand displays and advertising is banned at point of sale, a generic sign saying that tobacco is sold is allowed. This promotes availability. Tobacco retailers can also promote products to customers by word of mouth.

The final type of rules is independent of attempts to control marketing, and is about what form and content are required for warnings. The content may include facts about the adverse effects of tobacco use, benefits of quitting, and information about toxin levels (see Section 5.5). Here the aim is to discourage use or at least ensure that any continuing or new use occurs in the context of some information about the risks; that is, it provides cues to inhibit use. Warning and other risk-related information can be required on packets, at the point of sale, on any permitted advertisements, or in conjunction with any depiction of trademarks or commercial mention of products.

Tobacco industry controls are often about reducing cues to use tobacco, while tobacco use control efforts and information provision requirements directed at industry are about increasing cues to discourage use. For cues to use, the effect on behaviour is often conditioned such that they will stimulate tobacco use unless actively resisted. By contrast, cues to inhibit use are more likely to operate via conscious processing.

Evaluation of tobacco industry control is first about assessing compliance with the rules. This is unlikely to be an issue where the rules are to control obvious activities of small numbers of companies (e.g. compliance with labelling requirements), but can be an issue where there is more potential for avoidance (e.g. many potential actors or where the action is not so obvious; e.g. payment/avoidance of taxes). Evaluation is next about determining the effects of the rules. What is involved here varies as a function of whether the rules mandate some actions (e.g. warning labels, higher prices) or whether they mandate removing something (e.g. promotional cues to smoke) that would otherwise be there. In the former case, issues of reactions to the change need to be evaluated. In the latter, the extent

of previous response to the cues (or other things) removed must be known before the impact of their removal can be effectively evaluated. As noted above, it is necessary to monitor and evaluate any industry actions that might occur to reduce the impact of the rules on their businesses.

Tobacco use control

Tobacco use interventions are those targeted at tobacco users or potential users directly. They include rules about use, attempts to provide messages aimed at providing information and changing attitudes and beliefs, and programmes to deliver interventions that can facilitate appropriate behaviour change, or in the case of prevention interventions, effectively inoculate against uptake of any of addiction-level use.

Rules about use include policies to make various places smoke-free (see Section 5.2). Smoke-free rules are generally designed to protect non-smokers, although in doing so they have effects on smokers that need to be understood. Rules could also be about which products could be used, and by whom. However, where there are restrictions on use of products or who can use them, they are usually also codified as rules against selling such products (e.g. smokeless tobacco) or selling to particular individuals (e.g. minors), so these are best considered under industry control even when the parallel restrictions are imposed on individuals as well.

Provision of messages essentially relates to mass media campaigns, where the intent is to expose as many people as possible to the campaign (see Section 5.6). This may include campaigns to promote programmes. Campaigns are designed to inform people and to make the issue emotionally salient enough to stimulate appropriate action. One of the enduring challenges of tobacco control is that because the main adverse effects of smoking are not evident until after a long lag time, smokers do not experience any significant sense of the harm they are doing, and thus tend to underestimate its harmfulness (Slovic, 1998). There are extra issues to consider in the evaluation of prevention campaigns. Focussing on an issue increases awareness of it and may increase interest, which if unchecked could lead to increased experimental use. Designing prevention campaigns or programmes in ways that overcome this increased interest requires thought. There is evidence that some prevention campaigns, especially those emanating from tobacco companies, can have adverse effects (Wakefield *et al.,* 2006), presumably through the increased interest in the issue they engender.

Programmes to disseminate interventions include rules regulating cessation medications, provision of services, and subsidies to products or services (see Section 5.7). The kinds of products/services vary, including self-help resources, stop-smoking pharmaceuticals and coaching or advice programmes of various types. As noted earlier, this volume is not concerned with evaluating the efficacy of these products or services themselves, but on evaluation of their community-wide dissemination and use. Beyond this, there is interest in considering the effects of the existence of cessation services on the broader community. There is some evidence that awareness of the availability of quit-smoking programmes can stimulate quitting activity even among those who do not use the services (Ossip-Klein *et al.*, 1991).

Evaluation of tobacco use interventions should consider both their intended effects and incidental effects. They need to be informed by a sophisticated understanding of psychological principles, and where there are competing psychological processes involved, it is important to put in place measures of all relevant processes. Where additional effects to those sought are known (or hypothesised) they can become further justifications for action (or inaction, if they are or might be undesirable).

Use of logic models

Achieving a comprehensive approach to tobacco control requires adoption of a range of different strategies, underpinned by differing constructs and theories. It is important to spell out the relevant concepts to consider in each area in which a policy intervention might be

planned. The WG has adopted the strategy of encouraging the use of logic models or flow charts to spell out the main constructs that need to be measured for each type of policy. The criterion we adopted was to divide an area to the point where the causal pathways were sufficiently different to make dealing with the various possibilities difficult within the one frame. The WG used Figures 1.4 and 1.5 as generic models, but as will be seen, found the need to modify them considerably for some policy areas. We accept that as knowledge about how some of these interventions work accumulates, new distinctions may become necessary, which could lead to further subdivisions of intervention type. Further, in some cases, distinctions may be shown to be of lesser importance, allowing some of the existing boxes to be combined. It is only once a coherent theoretical model of the domain has been established that determining the constructs to measure becomes possible.

Measurement issues

Measurement is critical to evaluation. To measure the concepts of interest, these concepts must first be defined in ways that make them amenable to measurement. These definitions constitute the constructs. Constructs can be operationalised in many ways. This operationalisation must come from a clear consideration of the concepts and thus of the underlying theory. Because constructs are defined in terms of the theory and not directly in relationship to what measures them, error is localised in the imperfect relationship between the underlying construct and the measures used to assess it. Many of the concepts that need to be measured are not directly observable, or, where they are, they sometimes stretch the capacity of the respondent to recall or otherwise come up with a valid answer (e.g. remembering quit attempts months or years ago). As a result, most measures are subject to a range of possible biases as indicators of their target constructs. Exceptions are characteristics such as sex and date of birth, which in most cultures at least can be reported very reliably (although not in all). One of the great challenges of measurement is that the measures that are most easily obtained are often not ideal operationalisations of the constructs of interest. For self-reported data, most things people report are used as indicators of behaviour patterns or of underlying beliefs, behaviour patterns and/or understanding, not as simple answers to the question. The lack of direct measures also occurs for many physical measures. For example, cotinine levels are sometimes used to assess intake of nicotine or extent of smoking. However, because people differ both in size and in rate of nicotine metabolism, cotinine is a biased measure of intake or exposure at an individual level, although it can be a good estimator at a population level. The problem of the measure that is available not being a direct measure of the construct of interest may be greater when existing data are used, as compromises are commonly made in the interests of being able to use what is at hand. These data were often collected for quite different purposes to those of focal interest, and thus the measures used are often of related constructs, not the exact ones being studied. Dependent on the study, evaluators may be forced to use measures of constructs with different limitations. They need a language to help them talk about the quality of measures in relationship to the constructs they are using the measures to assess. Unfortunately there is no consistent language for talking about these distinctions, and the WG were unable to develop one for this volume. The WG views the development of such a language as critical to reducing the potential for conceptual confusion that can occur from failing to consider the limits of specific measures to actually measure the constructs evaluators are interested in measuring.

Determining what to measure

Choice of potential measures begins with an elaboration of the theory or theories as to how the intervention might work, including the range of expected outcomes and potentially mediating (or intermediate) and moderating variables (effect modifiers), as well as incidental effects. It might also

consider questions like: "What outcomes will lead to health gains?" and "What might influence policy adoption and/or continuation?" Evaluators should also consider whether the same outcomes are relevant to all cultures. For example, in Islamic countries and others where alcohol use is prohibited or not socially significant, consideration of smoking policies in bars is of little interest. Also the relevance of some issues can change as a function of a society's status in regards to tobacco control efforts. For example, support for and reports of smoke-free hospitals are now so high in many countries, it is no longer necessary to ask. However, in countries where passive smoking has not become an issue, asking about smoke-free hospitals may be critical to assessing emerging community concern. This analysis identifies the concepts that it would be desirable to measure.

Next, evaluators need to consider how they want to operationalize the concepts as constructs. This needs to be done in a way that ensures that the constructs are structurally independent of related constructs they might want to relate them to in causal pathways. Further, they need to consider whether the construct can always be measured in the same way. Physical measures typically measure the same thing regardless of context, but answers to questions may not. For example, the direction of social desirability biases might switch as smoking becomes less socially normative. For any given study, they must assess how well the constructs of interest can be measured. Where adequate measures do not exist, there will unavoidably be gaps in the modelling. Sometimes these gaps can be covered, at least in part, by using sets of measures of related constructs.

In Chapters 4 and 5 of this Handbook the WG provides guidance on measures that might be used in various evaluation contexts. For any domain of interest we attempt to characterise constructs that might be measured as one of:

1. *Core constructs:* those that should be included whenever this domain is being studied. These will include key outcomes along with major theorized mediators and moderators. Not having measures of any of these is likely to compromise the study, or at least limit the range of inferences that can be drawn.
2. Important *complementary constructs*, to use for detailed investigation of a domain.
3. *Other measures or indicators* that may add some limited or uncertain value, but which we cannot recommend (for or against), or only recommend in limited circumstances.
4. *Not recommended*: these only need to be specified for commonly used measures that have been shown to have no utility.

The quality or validity of the measures used for each construct also must be considered. *Validity* of measures refers to the extent to which they actually assess the construct they are designed to. This can be assessed through the relationship between the measure and a gold-standard measure (criterion validity), or by showing that the measure related to other theoretically related constructs as hypothesized (convergent validity). One form of convergent validity is predictive validity, where the measure is shown to predict outcomes as theorised. A valid measure of one construct is unlikely to be an equally valid measure of even a closely related construct. Also, the validity of a measure may vary as a function of how it is being used. Thus reports of awareness of environmental cues are not a valid measure of the extent to which any single individual is exposed (because of differences in sensitivity), but may be a valid measure of overall community exposure (as the individual errors are assumed to cancel out across the population). Validity also only relates to the contexts in which it is established. As the context changes the validity of a measure may vary. For example, self-reported age is generally a valid measure of how old somebody is. This is so in cultures where birthdays (anniversaries) are important occasions, but may be less so in cultures where people take no notice of birthdays. Also the validity of measures varies directly with the precision required of the measures: measures that may be valid for detecting large-scale effects might not be adequate for detecting small effects.

The WG uses the following broad categories to provide an

indication of the quality of measures:

- *Gold standard measure.* Established valid measure of a construct of interest that is better than alternatives in all ways.
- *Clearly validated outcome or predictor*. There is evidence that this is a good way of measuring the construct, in at least some specifiable contexts. Limits to validity should be noted.
- *Evidence of utility.* There exists some validity data, but it is not strong. It might be one of a range of alternatives with no clear way of differentiating between them. These should only be chosen when no better measure is available.
- *Face validity*. This involves an analysis of the extent to which the question taps the construct, and may be all that is available for single item self-report measures.

Where possible, we also provide an indication of the sensitivity of the construct to measurement error. For example, how robust is a question to differences in wording? Or indeed, might wording or contextualizing statements need to differ by context and/or by characteristics of the respondent? For example, some questions need to change for use with current smokers as compared to ex-smokers; e.g. "How confident are you that you will be able to stay quit, if/when you try (The last qualifying phase is not needed for ex-smokers)?" Users of this manual should keep in mind that the quality of a measure may be dependent on the type of study in which it is collected and the use to be made of it. The assessments made here assume the measures are made in appropriate circumstances.

Types of data used in evaluation

The type of data needed for evaluation varies, and in some cases it can be found in existing data collections, although sometimes measured in ways that are less than ideal for the new purposes to which it is going to be put. In some cases, measures of the variables of interest are available from more than one source. In these cases, decisions need to be made as to which sources of information are most useful. Issues to consider here are validity, practicality of collection, and the extent to which the data can be related to specific individuals. However, in most cases, the necessary information will need to be collected, giving the researcher greater control over the ways in which the relevant constructs are measured. Some of the main types of data and major ways of collecting it are outlined below.

1. Documentation of policies.
Critical to any form of evaluation is documenting the nature of the intervention. Documentation of policy can occur at two levels: the espoused intent or formal policy (something that is typically documented), and the actual program of activity that is put in place to implement it (which is usually more difficult to document). Policy documents should be collated and coded in ways that allow appropriate comparisons to be made. There is now an international repository of information about the content of national tobacco control policies (See Section 4.1), making this task easier, at least for national-level policies. Some countries collect this information for sub-national policies, but in most cases, the information will need to be collected from each jurisdiction. Where there are many such sets of rules (e.g. of workplaces, local governments), it is usually more convenient to either obtain samples of policies, or to use respondents in population studies to report on the rules that apply to them. Clearly, this latter form is subject to the problem that ordinary people often do not know about rules, and where they do not, may respond in terms of what they remember. For example, when asked if there are bans on smoking in their workplace, some will know the formal rules and respond appropriately, whereas others may know the rules but respond in terms of what actually happens (e.g. if there is a rule, but it is ignored, they will report that there is no rule, interpreting the question to mean, "Can people smoke?"). Others will only be able to answer in terms of what they infer from their recalled observations, e.g. "Nobody smokes there, so it must be banned." This means that such reports may not

be able to help differentiate between policy existence and policy implementation. Indeed, generally there are difficulties in directly determining implementation, especially for complex policies independent of their effects. This is only a problem when the research questions include asking whether problems with a policy occur at the level of policy content, or are a problem of implementation.

2. *Identifying changes in the environment or factors that might moderate policy effects.* The challenges of doing this differ by the environment under consideration.

a) Mass media. Monitoring of national and regional media, with sampling of communities for audit of local media, is the most objective source of what is potentially available. This does not cover some important sources like the Internet. An aggregated respondent report is useful where there are sufficient observations per community unit. Individual reports are subject to sensitivity bias, such that when thinking about quitting, or trying to quit, the person is likely to be sensitized to mentions or images of tobacco or smoking. This means that respondent reports should not be used as indicators of exposure in most individual-level analyses.

b) Physical environment. These consist of rules about public tobacco use and cues to tobacco use from things like point-of-sale displays, billboards, and posters. They can be collected through observation in sampled settings. They may also be estimated from reports from relevant organisations (e.g. of workplaces as to the restrictions on smoking), but are assessed more often by reports from ordinary citizens as to what they experience, or for smokers, what they actually did (e.g. "when last at a restaurant, did you smoke?"). These reports can be averaged across communities to estimate overall levels of these features. Like other respondent reports, these are subject to sensitivity bias, limiting their use for individual-level analyses.

c) Production and sales data. Various forms of sales data, or proxies for sales data, may be available, usually related to reporting on taxes and excises. These may be national-level, but in some cases can be separated by type of outlet or locality. At a national level, there are some international repositories of this information (see Section 4.2). Self-report of price paid is a fairly accurate indicator of prices, but little is known of possible systematic biases.

d) Characteristics of tobacco products on the market. These include composition and engineering features of products and performance characteristics. These can either be gathered from the manufacturers or through independent testing.

3. *Effects on and characteristics of individuals*

a) Self-report data. Characteristics of individuals (knowledge, attitudes and behaviour) are generally only available from self-reports (some scope for proxy reports, but limited beyond smoking status). Self-report data can be of internal cognitive states that are not independently verifiable (e.g. of attitudes, knowledge or experiences), as well as of things that can, at least in theory, be validated, such as behaviours. Sometimes answers to questions can also be used to infer internal states of which the respondent is either not aware or not thought able to report accurately (e.g. personality traits). Many countries have routine behavioural risk factor surveillance studies and/or tobacco specific surveillance studies, and these can be useful in a range of contexts. Many countries use standardised methods and questions, and are working towards common repositories of data (see Section 4.3). Self-reports are affected by question wording and by other aspects of the ways in which the information is collected (see Section 2.2 for some examples).

b) Physical measures. This includes biological and chemical measures (e.g. of cotinine levels). These are often used to measure behaviour indirectly, but this should be done with caution. Limitations of these measures as well as their

strengths are well documented (Benowitz, 1996a; Matt *et al*., 1999; Al Delaimy, 2002).

c) Proxy reports. For observable aspects of behaviour, reports of others who know the target individual may be useful.

Survey methods for evaluation

Survey methods are crucial to many forms of policy evaluation. These can range from surveys of individuals to surveys of informants about the activities of organisations (e.g. of governments or workplaces). Two key issues are addressed here: the sampling frame and the way the questions are asked and answered.

Sampling: To be able to generalise to a population, the sample needs to be representative of the population. This is a function of both the sampling frame and participation. It is thus desirable to have broadly representative samples, recognizing that true representativeness is unattainable. Participation is also crucial. Any biases in participation threaten representativeness. Because often nothing is known about all or some of those who do not participate, quantitative estimation of biases is either impossible, or partial at best, meaning their likely effects need to be inferred. The higher the response rate, the less likely major biases are, but unless the rates are close to 100%, biases can occur.

Sample size is another important consideration. The two main factors to consider here are the size of effects that are expected (or the required power to detect) and the desire to explore potential moderator effects. In principle, making a study larger does not improve its representativeness. However, because size does increase power to detect moderator effects, larger samples can be used to increase confidence in the generalisability of the findings to all groups who have a sufficient sample size for such possible interactions to be tested.

Question asking: The main issue with surveys is inconsistency and bias in the ways in which people respond to questions. This is part of a general phenomenon of the frame of reference or context for the question affecting how it is understood, and thus how it is responded to. Variation in frame of reference includes mode of surveying (e.g. face to face vs. phone interview vs. self-completion). There is emerging evidence that some modes of surveying result in better response rates for sub-sections of the population. There is an urgent need for research to develop optimal methods for calibrating both questions and sample characteristics across modes (see Dillman & Christian, 2005, for a discussion of general issues concerning mixed-mode surveying). As it is beyond the scope of this volume to document the entire range of issues corresponding to questions (there are several excellent texts on this topic; e.g. Foddy, 1993; Fowler, 2001), we deal only with two issues in this chapter. These are the time frame over which answers apply, and cultural factors in interpreting question meaning.

The time interval over which the response is deemed to be valid is a crucial issue in testing causal models. Causes precede effects, so one must assume that predictor variables when measured at the same time as outcomes, predated the occurrence of the outcomes. Sometimes questions are given a time frame or timing of events is asked for to assist in determining sequences. Self-reports of periods or of dates are subject to biases in reporting with events sometimes displaced in time. Self-reports are typically better for recent events (due to memory effects). Salient events may be reported as experienced more recently than in reality, and less salient events are prone to be forgotten.

Aside from issues concerning the context of survey delivery, the way in which respondents interpret questions and response formats affects their answers. One key aspect is the extent to which the conceptual framework underpinning the questions reasonably applies across the cultural contexts under consideration. As research moves from studying issues like tobacco within Western European and North American cultures, to studying tobacco use across cultural settings where there may be different values and assumptions, there is a need to question the underlying assumptions that frame the research. Within all cultures, there will be variation that researchers should try to characterise and understand. The possibility that cultural differences may compromise the

utility of some questions needs to be reviewed on a case-by-case basis. Some of these issues and methods for overcoming them are covered in Section 2.2.

In principle, the response to a question can be directly compared when the respondents are answering the same question. People generally assume this means the same wording. However, under some conditions, the same wording can result in quite different questions being answered, and different wording may be required to achieve equivalence. The most obvious example is asking questions in different languages, but it can occur for the same language where respondents' assumptions about what is being asked can vary systematically, and achieving equivalence requires different contextualising words for different individuals. This can be caused by words having different nuances in different cultures, or effects due to the familiarity and or normativeness of the issues being asked about.

As surveys become standardised, there is a tendency for surveys to converge on common ways of asking questions, thus implicitly operationalising the constructs they are interested in. To the extent that either the operationalisation has an arbitrary element or the measure is flawed, there is a risk of institutionalizing error. To avoid this, it may be important to analyze whether different ways of asking questions may improve the ability to measure a construct. There is always a role for asking questions in different ways. Where the answers are relatively invariant to the form of wording, one can have considerable confidence in generalisability across the inevitable wording differences between languages. However, where responses are sensitive to wording, it is less likely that different forms are actually measuring the same construct, and extra care will be required in translation.

Study designs for evaluating population interventions

To best understand the implications of policy change (including community-wide dissemination of interventions), research designs should be as strong as possible. In Section 2.1 the relative strengths of various evaluation designs are can-vassed. In short, evaluation is strengthened with more observations (both before and after the intervention) within the population an intervention occurs in, the more populations that are studied in parallel, and the more alternative explanations for outcomes that are assessed within each study. In addition, the use of cohorts adds considerable power by allowing mediation and moderation effects to be tested more precisely. Finally, representativeness of the sample to the study population can increase the generalisability of findings. The ITC study (Fong *et al.*, 2006a) is a good example of what can be achieved by attempting to implement as many of these attributes as possible.

Achieving the strongest possible evaluation involves putting in place measures of key outcomes (at least) as long as possible before the policies are implemented. Obviously the best way to do this is if the measures can be part of the country's ongoing surveillance system. Where this is not possible, the studies should be implemented as early in the process of discussing policy change as possible.

For detection of trends, it is important that both sampling frame and participation rates remain constant. This is to maximise the likelihood that biases are likely to remain constant so that any changes are unlikely to be due to a sampling effects. Repeatability is more important than representativeness for determination of trends because it requires comparability between estimates over time.

Such a research agenda requires monitoring of all relevant variables in a diverse range of communities or jurisdictions over a period of time in which there are differences in policy implementation between those communities. This will include use of repeated cross-sectional surveying, and where possible, more in-depth longitudinal cohort studies of samples of relevant individuals (e.g. smokers, and young people at risk of uptake), to begin to explore how the changes come about and whether some groups are affected differently to others. This surveying will need to be complemented by longitudinal monitoring of ecological variables. The level (nation, state, local area) of the variable measurement will determine the

practicality of maintaining ongoing monitoring of all activity or whether some sampling is necessary.

Such a program of data collection is needed to provide the infrastructure necessary for understanding the mechanisms of population level change. Among other things, it would increase understanding of which factors are culture-sensitive, and which are not, and how the roles of various factors change as a person's position towards changing and adopting target behaviour changes. Similarly, it would allow for an understanding of how community readiness to change affects realized change and how readiness can be modified, as well as the conditions that facilitate the institutionalization of change. For policy makers, it can provide information on need for further action.

Drawing conclusions about causes

The approach the WG has taken to evaluation shares more with the methods used in epidemiology to determine causes of illness, than the reliance on RCTs to assess clinical interventions. As a result, when considering criteria to use in drawing conclusions about the effectiveness of policy interventions, we have adapted the criteria used in the epidemiology of disease (Hill, 1965). The adapted criteria are:

- Magnitude of the observed effect, particularly in relationship to known naturally occurring variations;
- Temporal relationship between intervention and change in target outcome;
- Exposure-response gradient;
- Biopsychosocial plausibility; that is, the effects can be explained as occurring through a plausible mix of biological, psychological and/or social process;
- Coherence across lines of evidence with different threats to validity, e.g. similar results using aggregate data and self-reported consumption could rule out response biases;
- Coherence of results from demonstrations of effects on different parts of the theorised causal pathway, or by demonstrating efficacy of components (e.g. the evidence of efficacy of many cessation aids makes it more likely that they have effects when delivered as part of programmes of help);
- Evidence that this type of intervention can have effects on other comparable outcomes (e.g. on other behaviour patterns);
- Consistency of observed effects across studies and populations, or clear patterns in the variability to demonstrate limits to generalisability;
- To which we would add: Elimination of theoretically possible alternative mechanisms for explaining the observed effects.

Policy evaluation has added challenges to other forms of outcome evaluation, because policies usually occur in a mix and policies are only one set of factors that are responsible for the outcomes of interest. Smoking prevalence or rates of quitting are determined by multiple factors, and establishing the contribution of each individual intervention is difficult. The task of differentiating the contribution of all possible contributors to the observed effects is difficult.

In providing a summative evaluation of the effects of an intervention, we need to not only consider the size and nature of effects, we also need to consider the possibility that there is no meaningful effect. In particular, it is important to make a clear distinction between evidence of the absence of effects, and the situation where there is a lack of evidence; that we really do not know whether an intervention works or not. We recognize that science cannot prove the null hypothesis, but it can and should make statements about interventions where there is a consistent failure to find evidence of any meaningful effect.

We need to qualify effects with a statement about generalisability. Some interventions have similar effects in most contexts, others can be quite context-specific. This consideration needs to cover cultural adjustments to the intervention itself, as well as factors in the environment that might affect its potency (effect moderators). It is also important to consider the direction of effects. Some interventions might prove counter-productive. Clearly less evidence should be required to stop an intervention where the evidence suggests that it is counter-produc-

tive, than if it suggested no effect or only a small positive effect.

The levels of evidence framework used to evaluate discrete interventions is not appropriate for use in evaluating policy interventions. We see more promise in adapting the criteria used by the International Agency for Research on Cancer (IARC) for its Cancer Prevention Handbooks. This is essentially a four-level system: *Sufficient evidence* of an effect, *Limited evidence, Insufficient evidence*, and *Evidence suggesting lack of effect.* The WG's concerns with adapting this framework to our purposes, is that it does not allow for gradations in confidence of concluding no effects, it does not clearly differentiate adverse effects, and it does not consider issues of generalisability, all of which are desirable qualifiers in the policy context. One possibility would be to adopt a matrix as shown on this page, with additional statements on effect size (for established effects) and on generalisability.

The effect size could be rated as: Small, Medium, or Large (or undetermined). Consideration needs to be given to whether the highest level of certainly could be applied to interventions where there had not been a direct demonstration of effects on the target outcome, or whether inferred effects could ever be rated as better than Probable. For example, it has been shown that larger health warnings lead to more thought about quitting, and that more thoughts predict future quitting. However, nobody has shown that there is more quitting in the context of stronger health warnings being introduced. How reliably can one conclude that stronger health warnings stimulate quitting?

Finally, once the effectiveness of an intervention is established, less powerful research designs will be needed to monitor continuation of effects and/or to assess whether similar magnitudes of effect are attained with new populations. It is only when there is reason to believe that there are real differences that stronger research methods might need to be reapplied.

How to use this Handbook

This Handbook is designed as a guide for program and policy evaluators. The WG hopes it will be used as a tool for training new evaluators and those who need to understand evaluation principles. It can act as a reference source for arguments about the role of evaluation and the way to think about evaluation, and by extension the development of effective interventions. In doing so, we hope it provides a framework for increasing the scientific credibility of the field, by helping to show that policy evaluation has rigorous methods and can make important contributions to knowledge.

We also hope it will act as a stimulus for further action to improve evaluation methods and measures. As such, this Handbook will need to be kept as up-to-date as possible. This might involve periodic revisions once the principles have been tested, or some other mechanism for moving our expected standards forward. There is a particular need to update the material on specific measures and on the status of data repositories, as these are in a constant state of change.

We hope this Handbook will provide a stimulus to work towards greater coordination of the ways in which policy evaluation operates and the development and/or expansion of international repositories to collect the relevant data and reports, and user-friendly ways to extract this information and synthesise it.

Some future actions the WG would like to see:

- Work to coordinate and arrive at a set of core terms that are most useful for our field.
- Work on what the criteria for validation should be for the

The evidence matrix

No evidence is available			
Possible effect:	**Negative**	**Not meaningful**	**Positive**
Probable effect:	**Negative**	**Not meaningful**	**Positive**
Established effect:	**Negative**	**Not meaningful**	**Positive**

various kinds of measures used, and how that relates to the different types of measures.

- Development and agreement on use of prototype formats for reporting on frequently repeated interventions, such as mass media campaigns. This will facilitate their combination into meta-analytic studies, especially important for understanding where and when things work.

In conclusion, this volume should be thought of as an important step in a process, rather than as a static recipe book for evaluating tobacco control interventions. The methods described and the measures provided are the best available today. The principles outlined in this volume will persist, but those principles require that methods and measures be adapted to the changing world. The WG has built into this Handbook some guidelines for seeking out the latest methods and some guidance in assessing the need to move beyond the measures and methods described here. We believe that this dynamic but systematic approach is the best way to approach the future because it provides a framework that allows evidence to guide action both before and after programmes or policies are implemented.

2.1 The importance of design in the evaluation of tobacco control policies

Introduction

The goal of this section is to describe elements of research design for evaluation studies and how they can form the basis for stronger conclusions about the impact of policies. The groundwork for evidence-based medicine has come from painstaking evaluation studies of treatment options. It follows then that the foundation of an emerging evidence-based public health policy must begin with building a database from rigorous evaluation of public health policies. It should be noted that the elements of research design that we offer in the domain of population-level tobacco control can easily be applied in efforts to evaluate any population-level policy or intervention in public health. Just as surely as the laws of gravity operate in Mumbai as they do in Lyon, the principles of causality, and the methods employed to make more confident judgments about causal relations, are not constrained by location nor area of research.

This section does not offer a comprehensive review of evaluation research design. (see Cook & Campbell, 1979; Shadish *et al.,* 2002; Rossi *et al.*, 2003 for discussions of evaluation research, and Rootman *et al.*, 2001 for the evaluation of health interventions). We focus on *impact evaluation*, that is, whether the implemented policy led to desired outcome(s), rather than other forms of evaluation, such as *process evaluation* (e.g. identifying and evaluating the processes that led to the creation and/or the implementation of a policy).

More specifically, our aim is to highlight how the inclusion of specific features in the design of a policy evaluation study can lead to more concrete conclusions about the possible causal impact of that policy. This section focuses mostly on the structural aspects of research design. Good evaluation design involves the selection of appropriate measures of high validity and reliability. Guidelines and recommendations for such measures, across tobacco policy domains, are provided in other sections of this Handbook.

This section does not provide a review of the statistical analyses that are employed in evaluation studies. However, we do wish to point out one common misconception about the role of statistical methods in attempts to ascertain causality from data*: causality is to be found in the design, not in the statistics.* No statistical method, not even those whose name may imply some special status in this regard (e.g. *causal models*) can confirm causal direction. A structural equation model (with or without latent variables) that yields a significant coefficient for A→B cannot be used by itself to conclude that A causes B rather than B causes A. To do so would be to fall prey to the logical error of *affirming the consequent:*

Statement: If A causes B, then the A→B path will be statistically significant
Observation: The A→B path is statistically significant
False Conclusion: Then A causes B

The advantage of more advanced statistical techniques is that they can take into account characteristics of the data to yield a "better" estimate of the A→B path coefficient. For example, structural equation modeling with latent variables (Bollen, 1989; Hoyle, 1995; Kline, 2005) explicitly models the measurement error from multiple measures of a construct (latent variable), so that the resulting estimate of the relation between that latent variable and another variable is free of the measurement error

that would otherwise have biased the estimate[1]. However, this statistical method does not advance in any way the argument that A causes B rather than B causes A. In fact, a system of variables with paths going in one direction will yield exactly the same model fit as if that same system of variables had all the paths going in the opposite direction.

The key to advancing the quest for causality is to be found instead in the design of a study. Here we offer a review of the elements of the design of evaluation studies that will increase the confidence with which causal statements can be made between and among variables (e.g. whether a tobacco control policy had a desirable causal impact on behaviour).

In our review of research design features for the evaluation of tobacco control policies, we describe the framework of the International Tobacco Control Policy Evaluation Project (ITC Project), which incorporates a number of the design features that are discussed here (Fong *et al.*, 2006a; Thompson *et al.*, 2006).

The importance of pre-evaluation knowledge in the design of evaluation of policies

The planning and design of evaluation efforts should be the first step in the process of formulating and implementing a policy (or any kind of intervention). This suggestion is part of the recommendations for "best practices" that the US Centers for Disease Control and Prevention created for tobacco control programmes in 1999. They strongly recommended that 10% of the total budget for a comprehensive tobacco control programme be allocated for evaluation and surveillance efforts associated with the programme (1999a).The WHO EURO Working Group on Health Promotion Evaluation made a similar call for resources for proper evaluation (Rootman *et al.*, 2001).

Planning should first identify the constructs that are theorized to be affected by the policy being evaluated (i.e. outcome variables and mediators), as well as those that could influence the strength of the impact of policies on those outcome variables and mediators (i.e. moderators). The choices of which constructs to include in an evaluation study come from this process. This Handbook provides descriptions of the constructs, and their measures, for many of the Framework Convention on Tobacco Control (FCTC) policy domains.

Identification of other possible events that might act as confounding factors (e.g. other tobacco control policies being implemented and programmes in operation, tobacco industry initiatives) should also be addressed in the planning stage. Knowledge of possible confounders may allow additional variables to be measured or design features to be incorporated, so that the evaluation of the policy can explicitly take them into account.

Causality

Ultimately, the goal of scientific inquiry is to attempt to identify causal relationships. The concept of cause has challenged and vexed philosophers and scientists alike through the centuries. The seminal work of epidemiologists, such as Doll and Hill (1950,1954), Wynder and Graham (1950), and Levin *et al.* (1950), on the association between smoking and lung cancer, stimulated the thinking about identifying criteria that would be used in the determination of causality in epidemiology. This influential work was the basis of the US Surgeon General's Report of 1964, and was summarized in several articles including one by A. Bradford Hill (1965). We have adapted the original nine considerations of Hill, in assessing the strength of evidence, into seven criteria concerning the possible causal impact of a tobacco control policy:

- Consistency of observed associations across studies and populations
- Magnitude of the reported association
- Temporal relationship between intervention and change in target outcome
- Exposure-response gradient
- Biopsychosocial plausibility

[1]This assumes that the common variance of the multiple measures of the construct perfectly capture the latent variable that the measures are intended to capture.

- Coherence of results across other lines of evidence
- Evidence that this type of intervention can have effects on other comparable outcomes (e.g. other behaviour patterns).

From criteria for causality to research design: the framework of Cook and Campbell

Cook and Campbell's (1979) seminal treatise on the relationship between research design of a study and the strength with which a causal relationship might be ascertained, is our starting point for a discussion of how design features can be employed to evaluate the impact of population-level tobacco control policies.

Central to the Cook and Campbell framework is the concept of *validity*. Cook and Campbell defined four kinds of validity that are critical in assessing the validity of a causal statement: *construct validity*, *external validity, statistical conclusion validity,* and *internal validity.*

Construct validity refers to the extent in which a measure captures the construct that it is intended to assess. An issue that arises in considering construct validity is the method of measurement and whether there exists a close or distant relationship between those measurements and the construct. In the area of tobacco control, examples include: Is cotinine a valid measure of exposure to tobacco smoke? Is the Fagerstrom Test for Nicotine Dependence (Heatherton *et al.*, 1991) a valid measure of nicotine dependence? What are the most valid measures of perceived risk among smokers? These basic measurement issues must be dealt with in order for the validity of a causal inference to be addressed with any substance or meaning. Sections 3.1 to 3.3 of this Handbook review the construct validity of measures to assess the effectiveness of tobacco control policies.

External validity, also known as *ecological validity,* refers to the extent in which the conclusions of a given study are maintained across different persons, settings, treatments, and outcomes (Shadish *et al.*, 2002). External validity considers issues such as whether a phenomenon studied in a laboratory setting, often involving university undergraduates, will be obtained in a "real-world" environment, which includes individuals from the general population. However, in the public health realm, two issues of external validity (whether or not the issue is expressed in these terms) arise. First, there is the importance of sampling. In evaluating a tobacco control policy being implemented in a large and diverse population (e.g. in an entire country), probability sampling methods will provide the best assurance that the study sample will be representative of the population from which the sample has been drawn and to which the intended intervention is directed. To the extent that a sample deviates from a representative sample, the external validity may be correspondingly reduced; however, it should be noted that this conclusion is not automatic. It may be that the way in which a sample deviates from the population is not (strongly) associated with the variables being analyzed; thus, the net impact may not be as great as might have been expected.

Another way in which external validity applies to the evaluation of policies and interventions is in the distinction between efficacy and effectiveness (the former referring to a treatment effect in a controlled context, and the latter referring to the effect of that same treatment in a more "real world" setting). In general, effectiveness is lower than efficacy. Interventions originally developed and tested in highly controlled experimental settings are often not as effective when implemented in the real world. This necessitates changes in an intervention when brought into real world settings in order to maintain its effectiveness, as in the more controlled settings.

The two types of validity described above set the stage for the next two forms, which deal with the relationship between two variables and whether the measured association is indicative of a causal relationship. For simplicity, our discussion revolves around whether there is a causal relationship between two variables, although the logic applies to relationships among more complex sets of variables.

Statistical conclusion validity refers to whether there exists a statistical association between the two variables. Issues surrounding

the consideration of statistical conclusion validity include: statistical power, assumptions of the statistical tests being employed, the inflation of Type I error rates due to the conduct of multiple statistical tests, unreliability of measures, as well as the selection of "appropriate" covariates/control variables in estimating the relationship between the two variables. Though correlation is important and necessary, it is not sufficient to imply a relationship for causation, as captured in the dictum "correlation does not suffice to establish causation".

Internal validity refers to the extent to which the study's design is rigorous enough to support the conclusion that the statistical relationship between two variables is due, at least in part, to a causal relationship. Here we focus on issues of internal validity, as adding design features to a study (e.g. a control group) is largely prompted by the objective of increasing the internal validity of the study. The most relevant threats to internal validity in the evaluation of tobacco control policies are presented in Table 2.1.

Basic study designs and features

We now proceed to a description of aspects of an evaluation study, and make a distinction between study design and a study feature.

The *study design* is the structural aspect of an evaluation study, defined by three dimensions:

1. *Who* the study is collecting measurements from relative to the policy that is being evaluated. Some evaluation studies only measure the impact of the policy by collecting measurements from those who were exposed to the policy; other evaluation studies, however, measure the impact by also collecting parallel measurements from those who were NOT exposed to the policy.
2. *When* the measurements were collected relative to the policy's implementation. Some evaluation studies only collect measurements after the policy was implemented; others collect measurements both before and after the policy was implemented.
3. *How many* measurements are collected. Evaluation studies vary in the number of measurement time points, ranging from a pre-post design involving one pre-policy and one post-policy time point, to a time series design involving many measurements over time.

A further design parameter arises in evaluation studies involving more than one measurement over time; that is, whether those multiple measurements are obtained on the same individuals (the longitudinal or cohort design) or on different individuals (the repeat cross-sectional design).

In contrast, a *study feature* is a non-structural aspect of a study whose inclusion will enhance the ability to address threats to internal validity. One such feature is the inclusion of multiple measures within the domain of the policy that is being evaluated, toward the goal of achieving convergent validity (multiple measures of the same construct should be related to each other). For example, in a study of the impact of graphic warning labels, we would have greater confidence that there was a causal impact of the labels if, after being exposed to them, smokers were significantly more likely to: (1) self-report that the warnings made them think about the health risks of smoking, (2) more likely to call a quit line, and (3) more likely to cite the warnings as a reason for seeking assistance for quitting, than if only one of these measures was included in the study.

Another study feature is the inclusion of measures that are relevant to some other policy that is NOT being evaluated, as it is not changing in the study population toward the goal of establishing discriminant validity (i.e. measures of different constructs should NOT be so related to each other). In the policy evaluation context, measures of the non-changing policy should NOT show change that is comparable to that in measures of the policy under evaluation. In addition, inclusion of measures that will allow the testing of mediational models are designed to elucidate the causal pathways between the policy and an important outcome variable, such as a quit attempt. For example, in an evaluation study of graphic

AMBIGUOUS TEMPORAL PRECEDENCE: Lack of clarity about which variable occurred first may yield confusion about which variable is the cause and which is the effect.

- Cross-sectional survey data are particularly vulnerable to this threat.

SELECTION: Differences in respondent characteristics between groups that could also cause the observed effect.

- For example, observed differences between countries could be due to characteristics of the inhabitants rather than to differences in policies. Cross-sectional studies are particularly vulnerable to this threat.

CONCURRENT EVENT CONFOUNDING (HISTORY): Events occurring concurrently with treatment could cause the observed effect.

- For example, observed differences between countries could be due to other events or some other intervention (e.g. mass media campaign) rather than to differences in policies. This kind of confounding also includes activities of tobacco companies, which may be covert. These other events can cause the observed effect to seem stronger or weaker, positive or negative, compared to the policy/intervention's "true" effect. Concurrent event confounding could occur in longitudinal (cohort) studies, as well as in cross-sectional studies.

TEMPORAL TREND CONFOUNDING (MATURATION): Naturally occurring changes over time could be confused with a treatment effect.

- For example, trends over time occurring prior to the policy being evaluated, that are unrelated to the policy, could mimic the expected impact of policy or an adverse impact of policy (e.g. bar revenues dropping prior to the implementation of the policy could be the cause of a decrease in bar revenues observed after a smoke-free law compared to before the law).

ATTRITION: Loss of respondents to treatment or to measurement can produce artefactual effects if that loss is systematically correlated with conditions.

- Artefactual effects due to attrition can occur in cohort surveys of different groups (e.g. countries) where the attrition rate varies across the groups, and that attrition is linked to the outcome variable either directly or indirectly, via its linkage with an important predictor of that outcome variable. Related to attrition is non-respondent bias, in which non-respondents in an evaluation study could be differentially affected by the intervention (e.g. the very disadvantaged, who may be missed by both the intervention and its evaluation). Note that attrition effects in cohort surveys and selection effects in cross-sectional studies both involve biases in the sample that could lead to artefactual effects.

CONDITIONING (TESTING): Exposure to a test can affect scores on subsequent exposures to that test, an occurrence that can be confused with a treatment effect.

- An example of this threat is the presence of time-in-sample effects in cohort studies: participation in prior waves of a survey change the responses at the current wave (e.g. knowledge items, if repeated, can lead to observed higher levels of knowledge because of taking part in prior surveys).

Table 2.1 Selected Threats to Internal Validity and Examples

warnings, confidence that the introduction of graphic warning labels was responsible for an increase in quit line calls, rather than a mass media campaign, would be greater if there were measures included of the mass media campaign (e.g. recall measures of the campaign), and that these measures were not correlated with the likelihood of quit line calls.

In short, the internal validity of an evaluation study can be increased by including multiple measures of the policy, or other intervention, that is hypothesized to be responsible for the policy's impact, as well as measure(s) of other possible causes.

Designs for evaluation studies

In considering designs, we use the terminology of Cook and Campbell (Cook & Campbell,1979; Shadish *et al*., 2002) in which X stands for the treatment/policy that is being evaluated (e.g. introduction of graphic warning labels, increase in taxation, smoke-free legislation), and O stands for an observation (e.g. a survey data wave, quarterly report of cigarette consumption, or a set of data gathered by an air quality monitoring device).

Designs without control groups

The one-group posttest-only design:

In this design, the researcher has conducted one post-policy observation on some relevant unit of analysis. For instance, the unit could be human respondents to a survey, consumption figures from an economic database, or a venue at which the levels of respirable suspended particulates are being measured. The diagram of this design is as follows:

$$X \quad O_1$$

O_1 occurs after the policy X has been implemented.

In this post-only design, there is no sense of what the observations would have been in the absence of X ; therefore, this design alone is very poor. It does not defend against any of the threats to internal validity except ambiguity about temporal precedence. The history effects, and all threats associated with changes over time, are uncontrolled.

Given that none of the threats to internal validity are dealt with in this design, its value for evaluating policies, or interventions of any kind, is low. And yet it should be noted that the absence of a pre-test in this design often arises when the need for evaluation is recognized too late for a proper pre-test to be planned and implemented. This highlights the need for evaluation strategies to be established well before the intervention is applied, as discussed earlier.

In an effort to estimate the impact of X, researchers sometimes ask post-only respondents to recall their behaviour, opinions, or attitudes prior to X, or to make a judgment as to how X has affected them since. One should be cautious about the findings of studies relying solely on such strategies, as considerable experimental and survey evidence has demonstrated that such recall is subject to strong retrospective biases related to the respondent's theories on how the intervention might have affected them. These recall biases can occur when the respondent remembers the past as being more similar to the present than it actually was (consistency bias). When asked to estimate whether an intervention affected them, the recall bias could be in the direction of greater *contrast* (i.e. remembering the past as being more discrepant from the present than it actually was, with the magnitude of this contrast bias being correlated with the respondent's belief about the strength of the intervention (Conway & Ross, 1984; Ross, 1989; Pearson *et al*., 1992)).

Another more promising method of amplifying the value of the one-group posttest-only design is to incorporate data about pre-policy observations that are available from other sources. For example, if a new tobacco surveillance survey were created after a tobacco policy had been implemented, incorporating prevalence data from *other* surveillance surveys conducted prior to the policy would offer some comparison with a pre-policy measurement. The adequacy of this strategy would depend on the similarity between the two surveys (e.g. sampling,

method of measuring the outcome variable(s)).

The one-group pretest-posttest design:

This design adds a pre-policy observation to the previous design, and is denoted as follows:

$$O_1 \quad X \quad O_2$$

Here the addition of the pre-policy observation allows the computation of the difference score, $O_2 - O_1$, some portion of which might be causally attributable to the intervention X. The presence of an explicit measurement of the pre-post difference makes this far superior to the post-only design.

This design is considerably better than the one-group posttest only design. There is an explicit measurement prior to the policy that is not inferred or reliant on the validity of a respondent's memory or estimate of effect. The O_1 acts as a control against which the post-policy measurement O_2 can be assessed. In a repeat cross-sectional design, when O_1 and O_2 are taken from different samples in the same population, the control exists at the level of the group. In a cohort design, when O_1 and O_2 are measured from the same individuals, there is an additional level of power: each individual acts as their own control. Thus, response tendencies (e.g. the tendency to use the high end of a response scale, or to agree with survey questions (also known as acquiescence bias)) are controlled for at the individual level. This leads to greater statistical power, and the magnitude of this increased statistical power is a function of the extent to which individuals' responses at O_1 and O_2 are correlated.

Multiple pretest-multiple posttest design:

This design extends the single-group pretest-posttest design by the inclusion of additional pretest measurements and multiple posttest measurements within the group that received the policy/interventions, as in this example with 3 pretest and 3 posttest measurements:

$$O_1 \; O_2 \; O_3 \; X \; O_4 \; O_5 \; O_6$$

With many time point measurements, this design becomes a time series design. Variations within this multiple time point model include multiple pretest-single posttest and the single pretest-multiple posttest designs. These designs provide opportunities for assessing the impact of policies/interventions on the time related trends in the outcome variable that are unrelated to the policy, but which without knowledge or measurement of those trends, would bias the measurement of the policy's impact. When present, time related trends constitute an important confounding factor against which the effect of the policy must be evaluated. An example of the importance of taking into account these time related trends is presented later in this section.

In addition, designs with multiple measurements over time allow the evaluation of policies/interventions whose intensity varies over time, permitting the possibility of correlating intensity of intervention (e.g. measured by programme expenditures) with its corresponding impact. An example of this approach was used in studies evaluating the California Tobacco Control Programme, which distinguished between three time periods characterized by different levels of program intensity: pre-programme, early programme, and late program (Pierce *et al.*, 1998a).

Designs with a separate control group but with no pretest

Posttest-only design with non-equivalent groups:

In this design, a control group is added to the one-group posttest-only design. This design can be utilized if the evaluation process started too late to conduct a proper pretest measurement. If individuals were randomised to conditions, the groups would be "equivalent" on average, as randomisation equates groups with respect to all features of the individuals being measured. However, in the evaluation of national-level tobacco control policies, or in other cases where

the unit of intervention is a jurisdiction or organization, there is no possibility of randomisation, and hence, no possibility of equating groups[2]. The resulting design is the posttest-only design with nonequivalent groups:

$$X \quad O_1$$
$$\quad\; O_2$$

Case-control studies fall into this category, and often include various procedures to enhance the possibility of causal inferences, such as methods for matching the two nonequivalent groups. Issues surrounding these methods are well-identified in the epidemiological literature (Rothman & Greenland, 1998), but it should be noted that some of them, although possible with medical records among patient populations, may not be possible for implementation in evaluation studies of national-level policies.

Pretest-posttest designs with a control group:

This design is the basic "quasi-experiment" in which the pre-post measurement of the group that received the policy is compared to another group that did not receive the policy:

$$O_1 \quad X \quad O_2$$
$$O_3 \qquad\; O_4$$

The quasi-experimental design combines both elements that were used to enhance the internal validity of the one-group posttest design; added is a longitudinal component and a between-groups component. In this design, the critical starting point for an assessment of the causal impact of *X* is the construction of a multiple difference score; the change over time of the intervention group is compared to the change over time of the group that was not exposed to the intervention. The expectation, if the policy was effective, is that the pre-post difference in the policy group will be greater than the pre-post difference in the non-policy group.

The internal validity of the quasi-experimental design, although generally greater than the single group pre-post design, is dependent on the extent to which the non-policy group is similar to the policy group (e.g. similar levels of economic development, tobacco use prevalence). The greater the similarity, the more reasonable the comparison will be.

Randomisation to conditions is impossible in studies of policies. The strategy of strengthening an evaluation study via control groups depends on the selection of those control groups and their similarity. Various strategies can be used to enhance the selection of control groups that are objectively similar to the policy/intervention group on dimensions that matter (e.g. smoking prevalence, socio-economic status, similar levels of tobacco control intensity prior to the policy/intervention that is being evaluated in the study).

It would be more reasonable, for instance, to compare the impact of graphic warnings in Canada to a control group in the USA than to a control group in Bangladesh. It should be noted also that the "similarity" is not limited to the characteristics of the group. Relevant concurrent events should also be similar in the two countries. If, for example, the impact of graphic warnings in Canada were compared over time with a control group in the USA, but during that time between the pre- and post-policy measurements there was a large decrease in taxes in the USA, but not in Canada, the test of the graphic warnings would be confounded by the fact that the control group had changed in ways that would mimic the hypothesized impact of the warnings. Although the discrepancy of the difference scores would be consistent with the

[2] It should be noted that even in a fantasy world where people are actually randomly assigned to live in two different countries, one of which implemented a policy that the other did not, the randomisation would simply equate the personal characteristics of the respondents across the two groups. On average, the two countries would be populated by people who were equal on age, gender, age of initiation, number of past quit attempts, attitudes about the tobacco industry, etc. But left uncontrolled, would be the concurrent events that might occur along with the intervention that was being evaluated. The randomisation of people would offer no assistance for eliminating the possibility that observed differences between the two countries was due to differences in concurrent events. This demonstrates the limitations of randomised trials in the real world, even if such were possible.

conclusion that the graphic warnings had a desirable impact, the pattern of the data could also be explained by a significant unfavorable change in the difference score in the US control group due to the decrease in taxes.

This example points out that the structural features of the design endow an evaluation study with the *potential* for teasing apart possible alternative explanations, but that full realization of this potential is found in the selection of measures and analytic strategies that are designed to test for the *causal mechanisms* that underlie an observed difference between a policy group and a non-policy group. These strategies are described below in the section on mediation.

Threats to internal validity and methods for reduction

Having described some of the basic designs and strategies used in evaluation studies, we now proceed to a discussion of the threats to internal validity and methods for reducing them. As mentioned earlier, the rigor of an evaluation study is not only found in its design, but also in the features added to a study to enhance its power and internal validity. Examples are provided below.

Ambiguous temporal precedence:

A necessary, but not sufficient condition for causality is that a cause must precede the effect. The temporal priority condition provides challenges to cross-sectional studies by measuring possible causes and effects at the same point in time. It should be noted, however, that the temporal priority condition refers to the temporal ordering of the underlying constructs that are being measured, rather than the temporality of the data collection or observances per se.

In most cases, it is relatively simple to establish that the policy precedes a measurement. Even in a posttest-only design, temporal precedence is established: the measurement followed the implementation of the policy. However, because the key question is whether the evaluation measure changed as a result of the policy (i.e. whether the policy caused a change in the evaluation measure), the single measurement made in the posttest-only design is insufficient even as the temporal precedence condition is satisfied.

This discussion highlights the importance of multiple time point studies in assessing the causal impact of a policy/intervention, and is illustrated in greater detail below.

Selection: systematic differences over conditions in respondent characteristics that could also cause the observed effect:

Selection bias refers to the fact that individuals in different groups (e.g. different states, provinces, countries) are non-equivalent; that is, they could differ on dimensions that are correlated with the outcome measures used for the evaluation of the policies. Selection biases are difficult to identify and eliminate. Randomisation to conditions of an experiment is a powerful method for equalizing potential biases due to the non-equivalence of characteristics of individuals. However, randomisation is not possible in studies evaluating national-level tobacco control policies; therefore, selection bias in some form remains in all evaluation studies.

One approach to dealing with selection bias within a given evaluation study is to select control groups that are as similar as possible to the policy group. Thus, in evaluating the impact of policies in Canada, using the USA as a non-policy control group would be advantageous, as they are quite similar on many cultural and societal dimensions. If a policy in Canada were evaluated using, say, Kenya, as a control group, the inherent differences in the two countries would be much greater, leaving room for many more confounding factors.

A second approach is to measure differences between countries on constructs that might vary and act as possible confounding factors in the evaluation of policies. For example, in evaluating a policy in China compared to the USA, a possible confounder might be the fact that China is known to be a more collectivistic society, while the USA is a more individualistic

society. Knowing this difference, the evaluation study could add a measure of individualism-collectivism (Triandis & Gelfand, 1998), and correlate this variable with the policy-relevant variables in each country. If individualism-collectivism was uncorrelated with the policy-relevant variables, then this would suggest that, even though the two countries differed on this, it was not correlated with the policy and thus could not be a viable alternative explanation for observed policy impact.

The third approach considers multiple evaluation studies of the same policy in different settings and different times (i.e. of the overall consistency of the effects). This is adopted from one of Hill's criteria. If graphic warning labels are found to be effective in motivating individuals to quit smoking in Canada, Thailand, Venezuela, Brazil, and Belgium, then our confidence increases in making a general conclusion about the causal impact of graphic warning labels. Making general conclusions about policy impact will not and cannot occur on the basis of a single study, but rather after the consideration of multiple studies across multiple countries and time points. This principle is not limited to the evaluation of tobacco control policies.

It is worth noting that lack of consistency across studies provides an opportunity to examine what factors might be responsible for that variance. It may be that studies with weak designs yield different conclusions than those with stronger ones. In tobacco research, it has been shown that tobacco industry-funded studies of secondhand smoke are much more likely to conclude that it is not harmful, which is at odds with the very large number of non industry-funded studies concluding that secondhand smoke is harmful (Barnes & Bero, 1997,1998; for review, see Bero, 2005)

History: events occurring concurrently with treatment could cause the observed effect:

The internal validy of studies that evaluate the impact of policies over time, is threatened by events occurring concurrently with treatment/target policy which could cause the observed event. It is often the case that one treatment/policy intervention is implemented in conjunction with other policies/initiatives relevant to tobacco control. There are often other events, programmes, and interventions that are ongoing at the time of the policy that is being evaluated. Therefore, a major challenge is to estimate the impact of a specific policy in the field of other interventions that are ongoing simultaneously.

This is likely a common occurrence. If a government launches a comprehensive tobacco control programme, a frequent and recommended strategy would be to implement multiple policies and interventions. This comprehensive approach might include mass media campaigns, higher taxation, advertising/ promotion/-marketing restrictions, bans, increased resources for cessation programmes, and/or campaigns to raise awareness of existing cessation programmes.

For example, in 2003, countries of the European Union implemented new tobacco-use warnings, which were prominently displayed covering 30% of the package area. This corresponded with the minimal standard of warning labels under the Framework Convention on Tobacco Control (FCTC). The ITC Four Country Survey was launched in October 2002, in order to collect the pre-policy data for evaluating the impact of this enhancement of the warning labels. In May 2003, the second wave was conducted in the same manner as the first post-policy data collection.

By the time of the second survey, another important tobacco control policy had been put into action. In February 2003, the United Kingdom implemented a comprehensive ban on advertising and promotion of tobacco-related products, via billboards, magazines and newspapers, direct mail, domestic sponsorship (May 2003), website advertising and promotions, and exterior signs in store windows. This second policy complicated the quest for measuring the impact of the enhancement of the European Union's warning labels. Below, we outline an empirical strategy for distinguishing the effects of different interventions.

Factors that also influence the outcome measures of an evaluation study of a specific

tobacco control policy include activities of the tobacco industry, which are designed to reduce or neutralize the effect of tobacco control policies and programmes. Without consideration of these countermeasures (which could include explicit inclusion of industry activity variables), a policy evaluation study could lead to incorrect conclusions.

Although the importance of identifying and measuring the impact of tobacco industry activities cannot be over-emphasized, the impact of such activities will vary depending on the outcome measure. Broad, downstream outcome measures, such as prevalence rates, quit attempts, etc., are likely to be most strongly affected by tobacco industry activities. In contrast, more policy-specific outcome measures, such as label salience or the self-reported extent to which a smoker states that the warnings have made them think about the health risks of smoking, would be less likely to be influenced by industry activities. And here there is a trade-off: the measures of policy impact that are specific to that policy are less vulnerable to influence by tobacco industry counter-activity; as the measures become broader (e.g. going from label salience to perceptions of risk to intentions to quit to quit attempts), they are more vulnerable to impact from tobacco industry influences.

Maturation: naturally occurring changes over time could be confused with a treatment effect:

Typically, the term "maturation" refers to natural changes in individuals over time, such as changes that children undergo as they grow older. However, the concept might instead be called "time-dependent changes that are unrelated to the treatment." An example of how this concept must be identified and controlled for, comes from the claim made by opponents to the comprehensive smoke-free legislation in Ireland that sales volume in pubs had declined as measured before and after the March 29, 2004 ban (Figure 2.1).

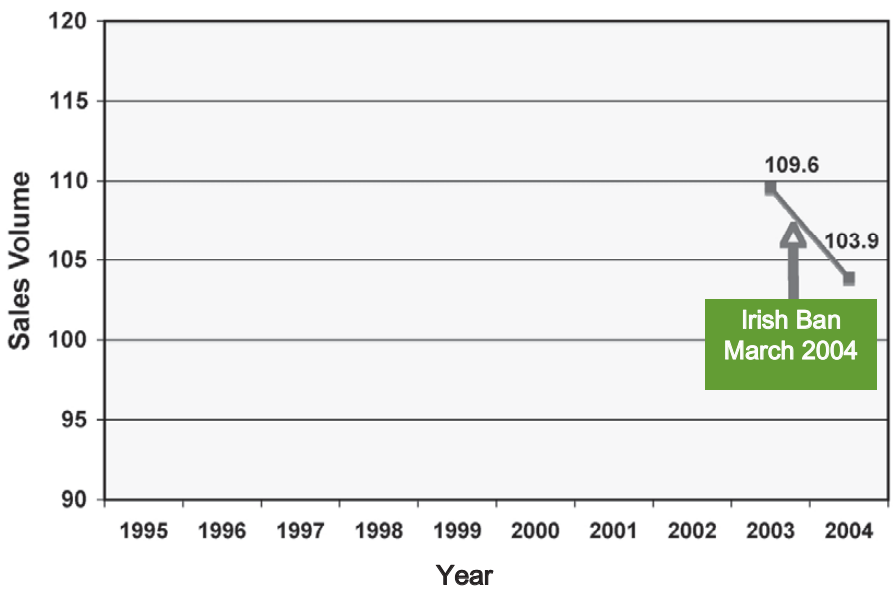


Figure 2.1 Pub sales volumes immediately before and after implementation of the Irish smoking ban in 2004

Source: Central Statistics Office of Ireland

Sales volumes are indexed so that sales volume in 1995 = 100

The data on the volume of pub sales before 2003 and after the 2004 ban, as shown in Figure 2.1, reveals that the volume of pub sales (indexed at 100 for volume of pub sales in 1995) in 2004 was lower (103.9) than it was for 2003 (109.6). With just those two data points, it might be concluded that the Irish ban caused a decline in sales in pubs.

However, Figure 2.2 presents the volume of pub sales for nine years (1995–2003) prior to the Irish ban. Taking into consideration the data from years prior to 2003 leads to a very different conclusion.

Sales volumes had been rising steadily since 1995, hit their peak in 2001, and then began to fall fairly steeply. When the full nine year profile is considered, the decrease between 2003 and 2004 does not appear to be any different than what would be expected by the secular trends. The decline between 2003 and 2004 was not significantly more dramatic than the declines experienced between 2001 and 2002, and between 2002 and 2003. When the more long-term "maturation" trends are considered, there was no greater decline after the smoke-free law had been implemented. Thus, the hypothesis that the Irish ban had a detrimental impact on the volume of pub sales is not supported.

Time trends can also work in the opposite direction. Suppose that the ban in Ireland was implemented between 1997 and 1998. If the evaluation study had been conducted with data from only those years, it would have shown an increase in sales, which might lead to the false conclusion that the ban was the cause of this increase. Again, consideration of the pre-policy time trends would reveal that the secular trend was indicative of increasing sales, and taking that trend into account would likely lead to a more proper conclusion that the ban had no impact on sales.

The implications for research design are clear: evaluating the

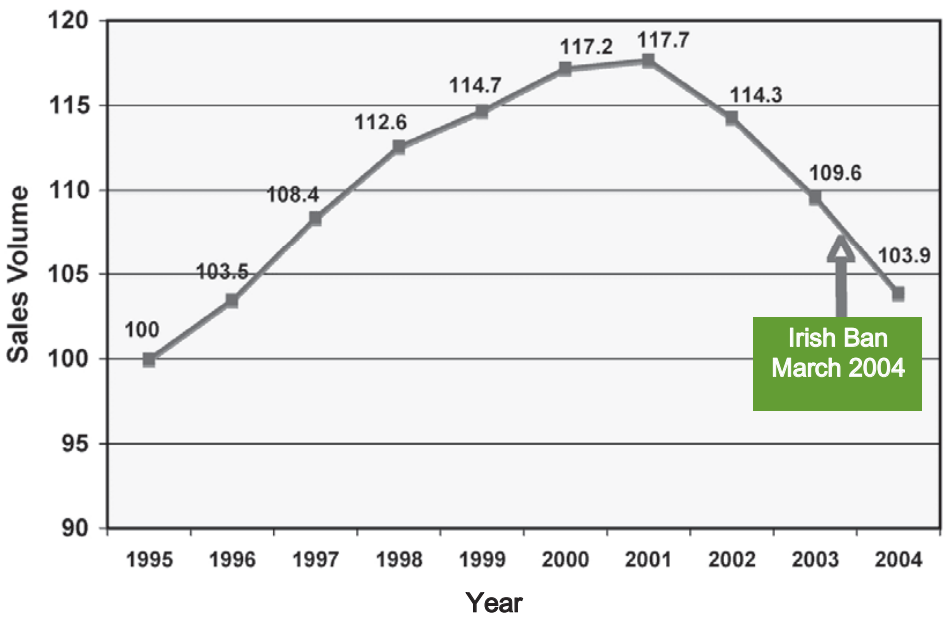


Figure 2.2 Pub sales in volumes in Ireland for the period 1995-2004
Source: Central Statistics Office of Ireland
Sales volumes are indexed so that sales volume in 1995 = 100

impact of policies is best conducted with the inclusion of data that allow the evaluation to take place within the context of time trends. This example highlights the value of having a surveillance system in place for collecting data over time on outcome variables of interest.

Although the Irish pub data illuminate the importance of time trend data, it also provides an example of how even good time trend data alone can sometimes be incapable of yielding a clear estimate of policy impact. To illustrate this, suppose the ban occurred in 2001 instead of 2003, and the evaluation was conducted with pub volume data from just 2001 and 2002. Here, consideration of the time trend might be taken to mean that the ban definitely reduced sales; however, it was still positive up to that point.

If only the time trend were taken into account, one might be even more confident of the conclusion that the ban decreased sales. However, in 2001, Ireland passed a law that limited the use of alcohol, which had an adverse impact on sales volume. Because of the presence of this known negative causal factor, the impact of the Irish smoking ban would remain ambiguous. Although time trend data are important in resolving some threats to internal validity, they fail to eliminate the threat to validity represented by concurrent events in the absence of information on the impact of such events.

A research design that is also concerned with understanding the impact of an intervention over time is the *interrupted time series* design (a specific version of this general design is the *regression discontinuity* design). In these designs, which require a fairly lengthy series of observations over time, the impact of an intervention can be measured by its impact on the mean function of the time series. In the regression discontinuity analytic framework, a distinction is made between the regression line that fits the data points (capturing the relation between the outcome variable and time) before the intervention, and the regression line that fits the data points after the intervention. The analysis compares the two lines; the effect of the intervention is measured as the difference in the slope, the intercept, or both parameters of the line. This kind of design can provide powerful evidence for the impact of a policy in its temporal context. There are a number of sources that describe these models (Trochim, 1984; Trochim *et al.*, 1991; Box *et al.*, 1994).

Time series approaches have been used in evaluating the impact of tobacco control programmes. For example, Pierce *et al.* (1998a) used piecewise regression analysis on time series data on cigarette consumption from 1983-1997 in California, versus the rest of the USA, to demonstrate that the California Tobacco Control Programme, initiated in 1989, led to declines in consumption. They also found that the impact of the programme was greater for the first five years than for the subsequent three. Biener *et al.* (2000) used similar methods to analyze prevalence data in Massachusetts versus the remaining US states (except California because of their similar comprehensive programme), and concluded that the Massachusetts programme led to a continued downward trend in prevalence, compared to the flattening of the downward trend in the other US states during that same time period.

Keeler and colleagues (1993) examined monthly time series data from 1980 to 1990 in California in their analysis of the association of cigarette prices, taxes, income, and anti-smoking regulations with cigarette consumption. Reduced consumption was found to be associated with tobacco control policies. They highlighted the impact of the tax increase in 1989, which led to a greater decline in consumption, followed by additional tax increases at other points along the time series.

In general, multiple time point data, particularly if such data are also available with control groups, provide strong potential for teasing out possible confounding due to time related alternative factors, and for providing confirmatory evidence for the impact of policies and programes. The strength of this potential (and therefore confidence in attributing changes in behaviour or some other important outcome measure) grows with the number of post-intervention data points, which means that more definitive conclusions might be reached

only after a greater delay than would be desired. The ability to come to more definitive conclusions increases with the number of other evaluation studies of a particular policy, or type of policy; within a specific (well-designed) study, the ability grows with the passage of time. Both require greater effort/time than is possible within a single pre-post evaluation study.

Attrition: loss of respondents to treatment or to measurement can produce artefactual effects if that loss is systematically correlated with conditions:

Attrition is a major concern in cohort surveys. In surveys about smoking, for example, those who quit are less likely to stay in the survey, even when specific provisions have been made for those who quit to move to a non-smoker/quitter survey, as in the ITC Surveys (Thompson *et al.*, 2006). Thus, it may be that if a policy or intervention is successful in increasing the proportion of individuals who quit, the greater attrition rate in the policy group, skewed as it is for those that quit, will attenuate the observed treatment effect (i.e. it will make the statistical test of group differences more conservative). Another potential bias due to attrition is seen in respondents with low socioeconomic status (SES), who are more likely to drop out. If the policy/intervention is more likely to have an impact on high SES individuals, the differential drop out will lead to an artificial enhancement of the treatment effect. The cumulative result of attrition will be the net effect of conservative and liberal biases, which will lead to uncertainty regarding the overall impact of differential attrition in any given survey situation.

Although attrition is unique to cohort surveys, non-response bias is a problem in cross-sectional studies, as well as cohort surveys. Non-response bias occurs when the surveyed sample differs from the population, because some types of respondents are less likely to agree to participate in the survey, or are less apt to be contacted in the first place. This poses the same problems as attrition; many factors contributing to non-response bias are present in biases from attrition.

As with all threats to validity, an approach to dealing with attrition is to measure its impact. The goal is to develop a model of the correlates of attrition that identifies variables that are associated with the likelihood of attrition and the strength of the relationship. Toward this end, it is valuable in cohort designs to replenish cohort members lost to attrition at each stage with newly recruited respondents from the same sampling frame. Differences between the responses of the cohort and the newly recruited replenishment sample can then be attributed to biases in attrition, and to time-in-sample effects, to which we turn next.

Time-in-sample: exposure to a test can affect scores on subsequent exposures to that test, an occurrence that can be confused with a treatment effect:

A *time-in-sample effect* (also known as *rotation group bias*) is a phenomenon whereby an individual's responses to the same question over time varies as a function of how many times the individual has responded to the same question in the past (i.e. the number of prior survey waves the individual has participated in (Duncan & Kalton, 1987)). In a cohort survey of nutrition, respondents were systematically rotated out of the survey, so that at each survey wave there were respondents who had participated 1, 2, 3, and up to 9 times before. It was found that respondents reported eating smaller quantities of food purely as a function of the number of prior survey waves they had been administered (Nusser *et al.,* 1996). It is valuable to take into account the time-in-sample effect in the analysis of cohort data.

Additive and interactive effects of threats to internal validity: the impact of a threat can be added to that of another threat or may depend on the level of another threat:

This statement reminds us that, as with any study, there exists more than one threat to internal validity and more than one source of bias in the estimate of an intervention effect. Some of these biases may

be in the direction of overestimating the effect; others may be in the direction of underestimating the effect. The impact of one source of bias can depend on the level of a second source of bias. For example, the overall impact of participation bias over time will depend on the level of attrition.

Cost effectiveness in the design of evaluation studies

On some dimensions, study design can be guided by a calculation of costs in relation to its benefits. The allocation of total sample size to number of clusters, and number of individuals within clusters, is one example where prior information (e.g. the incremental cost of conducting the study in an additional cluster; the intraclass correlation, a measure of the correlation of individuals within a cluster compared to the correlation of individuals belonging to different clusters) can be entered into formulas to create the "optimal" sampling design given specific resources available for the study.

In principle, the same is true for designing an evaluation study to reduce threats to internal validity, that is, a study that stands to yield a more confident judgment about the causal impact of the policy/intervention. But here, however, the process cannot be guided by formula or algorithm in the same way as can be accomplished in creating an optimal sampling plan. The increment in internal validity due to the addition of a second or third post-policy time point, for example, cannot be measured quantitatively. The reason is that the actual value is dependent on knowledge of the impact of spurious causal factors. The value of the second or third time point depends on whether the other causal factors would have exerted a policy-consistent or policy-inconsistent impact, which is unknown. In fact, if we actually felt confident enough about the impact of the other causal factors to put them in such a formula, there would be little need to actually conduct the evaluation study in the first place! Even though we cannot be specific about the value of a certain design feature in an evaluation study, we can make some general statements about the likely relative value of one feature or design element over another.

As described earlier, the single-group post-only design is not sufficient for evaluation of a policy (or any other intervention). So what could be added to this single measurement? There are two basic possibilities: (1) create a one-group pretest-posttest design by adding a pre-policy measurement from the same sampling frame as the post-policy measurement: either the same individuals who will be measured at post-policy (cohort design) or other individuals (repeat cross-sectional design); and (2) create a posttest-only design with nonequivalent groups by adding a post-policy measurement from another group who is not receiving the policy/intervention.

For example, suppose a researcher is planning an evaluation of the graphic warning labels introduced in Thailand in 2005 knowing that a post-policy measurement is required. But when adding another group to the design, should this second group be a pre-policy measurement in Thailand, or a post-policy measurement in another country, such as the neighboring country of Malaysia? It is strongly recommended that a pre-intervention measurement be added. This is because the starting point for all considerations of measuring the causal impact of an intervention is in the difference between pre- and post-policy (i.e. how respondents changed from pre- to post-policy on a label-relevant variable). Having an explicit measurement of this pre-post difference is much preferred to adding a control group (Malaysia), as the researcher would still have to infer what the outcome variable would look like in the absence of the policy at a time prior to the policy's implementation. As long as there is sufficient time to collect pre-policy data, this recommendation is also the easiest to implement. In the evaluation of national-level policies, it is simpler to obtain multiple measurements within one's own country than it is to obtain the same measurements in a different country.

Thus, the single expansion would favor the addition of pre-policy measures. In addition, the logistics of setting up the parallel study (e.g. a survey) in another country, with the establishment of a second research team, and the challenges of making the two parallel research efforts com-

parable in method and measures, would be great.

Summary of study design considerations

To summarize, in the absence of a randomised trial, there are two study design strategies that can be employed for the rigorous evaluation of the effects of policies. First is the use of measurements both before and after the policy's implementation. These measurements can be taken from either units (usually, but not limited to, individuals; the same logic would apply if the measures were of households, schools, or other venues) that are either the same (as in a cohort design) or different, but drawn from the same sampling process (as in a repeat cross-sectional design). The second design strategy is the use of a quasi-experimental design, in which one group that is exposed to a policy is compared to a similar unexposed group, as discussed above. Combining these two strategies in a single study yields a two-group, pre-post design, which offers a higher degree of internal validity than either feature alone. The utility of longitudinal designs is strengthened if there are multiple data collections before and/or after policy implementation, allowing more precise specification of effects (e.g. taking into account temporal trends that were occurring before the implementation of the policy).

Considerations of study features in the evaluation of policies

We have made a distinction between study designs and study features. In addition to the two design considerations, there are two study feature strategies that contribute to increasing an evaluation study's internal validity. The first is the measurement of policy-specific variables that are theorised to be affected initially after the policy is implemented. For example, in evaluating the impact of a new warning label policy on behaviour, one might reasonably predict that for the policy to exert its effect on behaviour, the target population must first report noticing the new warning labels (Hammond *et al.*, 2006). A second strategy is the measurement of policy-specific variables for policies that have not changed; such variables act as another form of control. In a country where labels have been enhanced and where taxation has not, for example, we would expect that label salience would be improved over time, but taxation-relevant variables (e.g. perceived cost of cigarettes) would not. Recommendations for measures in each FCTC policy domain are provided in other sections of this Handbook.

Combining the two design and two study feature strategies, along with the inclusion of other explanatory variables (covariates) that might help explain differences between two jurisdictions, creates a powerful research design allowing more confident inferences to be made about the causal effects of policies and/or combinations of policies. We now turn to an illustration of the use of these strategies in the International Tobacco Control Policy Evaluation Project.

The International Tobacco Control Policy Evaluation Project (ITC Project)

The ITC Project was established with the goal of measuring the psychosocial and behavioural impact of key policies of the FCTC on tobacco use among adult smokers (Fong *et al.*, 2006a; Thompson *et al.*, 2006). As smokers are directly affected by tobacco control policies, this understanding is crucial to assessing the extent to which the FCTC objectives are met, and of desirable and undesirable collateral effects. The ITC Surveys were explicitly shaped by the four strategies described above. To date (as of December 2007), the ITC Surveys are a set of parallel prospective cohort surveys of representative samples of adult smokers in 15 countries—Canada, USA, UK, Australia, Ireland, Thailand, Malaysia, South Korea, Mexico, Uruguay, France, Germany, The Netherlands, New Zealand, and China, with additional ITC Surveys under development in other countries (Bangladesh, India and Bhutan).

With these additions, the ITC project will be conducting

evaluation of FCT policies in countries inhabitated by over 50% of the world populations, 60% of the world smokers, and 70% of the world's tobacco users.

The ITC evaluation framework utilises multiple country controls, a longitudinal design, and a pre-specified, theory-driven conceptual model to test hypotheses about the anticipated effects of specific policies.

Conceptual model of the ITC Project:

The first step in creating the ITC Surveys was to determine how policies may achieve their desirable effects. How do policies work?

In order to address this important issue, a couple of assumptions need to be described. The first is that the most appropriate level of analysis, to understand the mechanisms by which policies may ultimately change public health outcomes, is that of the individual person. It is the individual who smokes or does not smoke, the individual who is influenced by anti-smoking media campaigns or by marketing campaigns of the tobacco industry, the individual who is or is not influenced by societal norms or by influences from close friends and family, and the individual who does or does not form intentions to quit and then either does or does not engage in an attempt to quit.

Having said this does not preclude the possibility, indeed the reality, that the individual can be influenced by forces at broader levels of analysis (e.g. social structure and organization), and by factors at even finer levels of analysis (e.g. individual differences of genetic susceptibility, such as high versus low metabolism for nicotine). Ultimately, however, it is individuals whose behaviour will or will not be influenced by policies, and in order for us to understand these behaviours, we must focus on the individual.

The second assumption is that there exists a causal chain of changes within the individual *through which the impact of policy flows*. This assumption directly relates to the idea of mediation: that policy causes changes in one or more constructs, and/or a chain of constructs within the individual, which then eventuates in behavioural change. The ITC Project team created a conceptual model of how tobacco control policies might work based on a combination of existing models from the psychosocial literature and from health communication theories. The resulting conceptual model, which is presented in Figure 2.3, guided the selection of questions included in all ITC Surveys.

The ITC conceptual model assumes that each policy ultimately has an influence on behaviour through a specific causal chain of psychological events. It is a general framework for thinking about policies and their effects on a broad array of important psychosocial and behavioural variables, and for testing how policy distinctions relate to their effectiveness. Several key characteristics of this conceptual model require further explanation. First, the model focuses on how policies affect the behaviour of individual smokers, and thus circumvents the potential hazards of making inferences about individuals from aggregates (i.e. policy studies in which countries are the unit of analysis, or individual-level studies that are repeat cross-sectional analyses conducted over time).The presence of macro-level causal forces that exert pressure on an individual, are acknowledged in the ITC conceptual model. For example, societal norms toward smoking, economic conditions, messages from the media that are either pro- or anti-tobacco use, and the influence of family and friends are taken into consideration. The model specifies, however, that the impact of those macro-level causes must be measured at the level of the individual through their perceptions of the presence of such factors (e.g. beliefs about the norms and expectations of society, close friends, and family on smoking). In the end, it is the individual who takes up smoking, who increases or decreases tobacco consumption, who does or does not attempt to quit, who is successful or unsuccessful at attempting to quit, and who may contract a smoking-related disease and die. Of critical importance, and a focus in the ITC conceptual model, is to capture and measure the influences of the many macro-level causes as

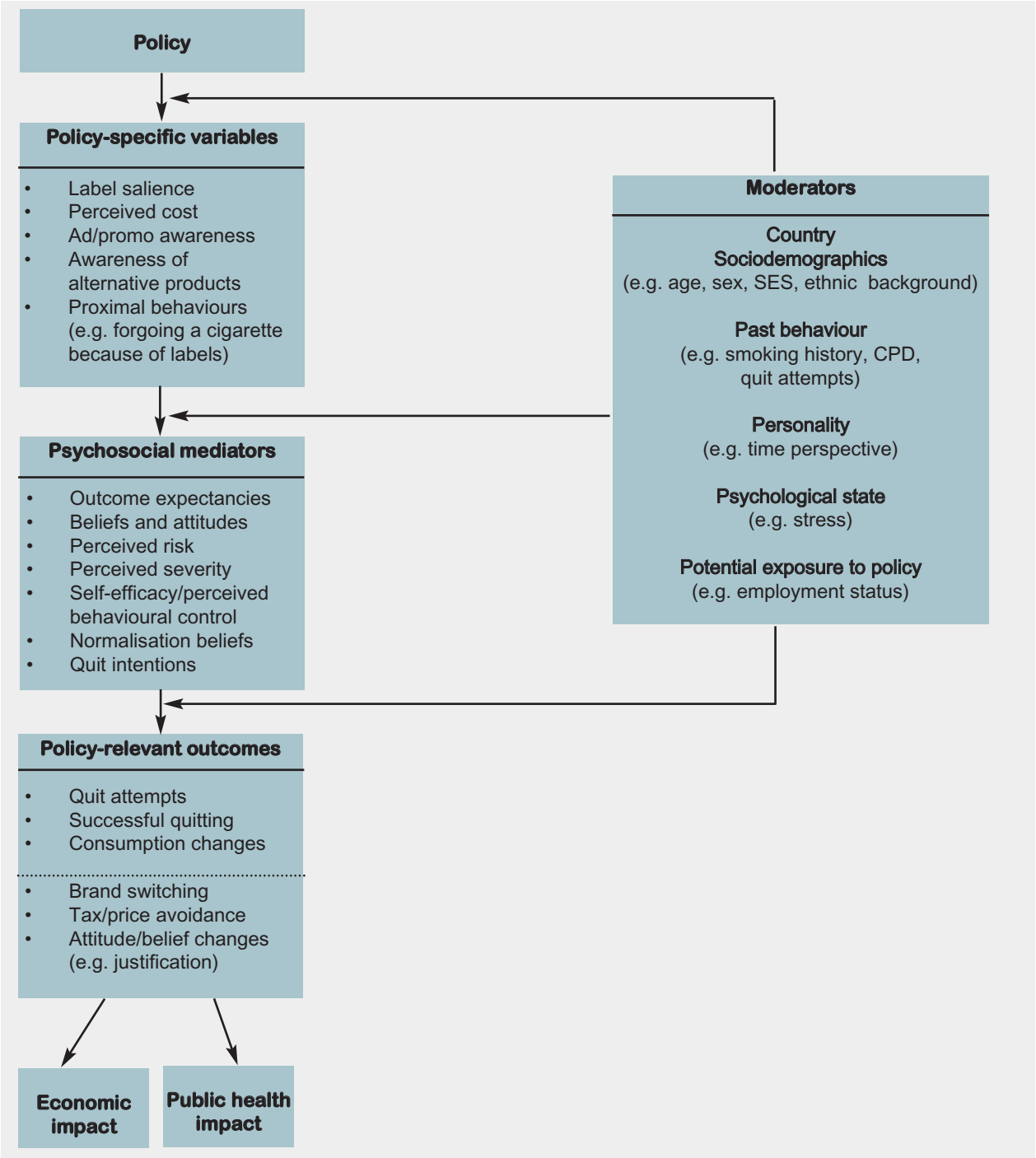


Figure 2.3 Conceptual model guiding the formulation of questions in the ITC Surveys
Adapted from Fong *et al.*, (2006a)

experienced by the individual. Ultimately, in order for us to understand the impact of policies and other macro-level influences on populations, it is essential to measure them at the individual level. It is a fallacy that the presence of macro-level causal forces requires that macro-level modelling be conducted.

Second, policies are seen as potentially affecting individuals along a variety of psychosocial and behavioural variables, of which there are two classes. The most immediate effects are those on the *policy-specific variables* (those variables that are proximal (conceptually closest), or most specifically related to the policy itself). Thus, new graphic warning labels should increase salience and the ability to notice warnings; price should affect perceived costs of cigarettes (for example, belief that cigarettes have become too expensive); and lifting of restrictions on alternative nicotine products should lead to increased awareness of the availability of those products. These effects may also increase the likelihood of discrete behaviours specifically linked to the manifestations of the policy such as smokers hesitating, or even forgoing or stubbing out cigarettes because of the warning labels. Examples of survey questions designed to measure policy-specific variables are presented in Table 2.2. Other sections of this Handbook describe these and other measures of policy-specific variables in each of the FCTC policy domains.

The more downstream effects are on the non-specific *psychosocial mediators*, which are conceptually distant from the policy and theorised to be affected by multiple influences, not just policies. Among these are variables such as self-efficacy and intentions, which come from well-known psychosocial models of health behaviour, including the theory of planned behaviour (Ajzen, 1991), social cognitive theory (Bandura, 1986), the Health Belief Model (Becker, 1974), and Protection Motivation Theory (Rogers & Prentice-Dunn, 1997). The ITC conceptual model holds that policies will affect these general mediating variables indirectly, through their prior effects on the policy-specific variables. As each policy has its own policy-specific variables, there exists potential to estimate the relative contributions of various policies to the outcomes of interest.

Third, the ITC conceptual model explicitly identifies the mediators of policy and articulates the goal of understanding the psychosocial processes that explain how and why a given policy may lead to changes in smoking behaviour. The longitudinal design allows the explicit testing of the causal chain of effects that is depicted in the model. With a repeat cross-sectional design, the capabilities of modeling the dependence of change in an outcome on the changes in an explanatory variable are diminished as data on the same individuals are not collected prospectively.

The policy-relevant outcomes that are measured in the ITC surveys include those that confer public health benefits (for example, quitting), but also include important compensatory behaviours that the smoker may engage in that, although responsive to the policy, may not lead to the economic and public health benefits that are ultimately the goal of such policies. For example, smokers may switch to discount brands in response to price increases, which would confer no public health benefit. The ITC Project thus attempts to provide a more complete account of the effects that may result from the implementation of a tobacco control policy, and includes both the detection of desirable effects and of unintended, undesirable side effects.

In summary, the ITC conceptual model is a causal chain model, and, as such, suggests that the policy-specific variables play a critical mediating role because they reside between the policy and the outcome variables that are important in public health (e.g. quitting behaviour). These causal paths, from policy-specific variables to behaviour, could be direct, but more typically will be through the more general mediators. In some cases, there may be pathways through several kinds of mediators, both the policy-specific, proximal variables, and the more general, distal variables. Policies are theorized to vary in the psychosocial ‘‘routes’’ that they take to affect behaviour, that is, each policy has a different

Policy Domain	Examples of Questions Measuring Policy-Specific Variables
Warning Labels	In the last month, how often, if at all, have you noticed warning labels on cigarette packages? Warning labels make me think about the health risks of smoking (level of agreement or disagreement with this statement)
Smoke-Free Legislation	Which of the following best describes the rules about smoking in drinking establishments, bars, and pubs where you live? – Smoking is not allowed in any indoor area – Smoking is allowed only in some indoor areas – There are no rules or restrictions For each of the following public places, please tell me if you think smoking should be allowed in all indoor areas, in some indoor areas, or not allowed indoors at all? – Hospitals – Workplaces – Drinking establishments (e.g. pubs/bars) – Restaurants and cafés
Price/Taxation	Where did you last buy cigarettes for yourself? How much did you pay for your cigarettes? The last time you bought cigarettes for yourself, did you buy them by the carton, the pack, or as single cigarettes? The last time you bought cigarettes or tobacco for yourself, did you use any coupons or discounts to get a special price?
Pro-Tobacco Advertising	In the last 6 months...how often have you noticed things that promote smoking? In the last 6 months, have you noticed cigarettes or other tobacco products being advertised in any of the following places: television, radio, at the cinema/movie theatre before or after the film/movie, on posters or billboards, in newspapers or magazines, on shop/store windows or inside shops/stores where you buy tobacco? Now I would like you to think about advertising or information that talks about the dangers of smoking, or encourages quitting. In the last 6 months, how often, if at all, have you noticed such advertising or information?
Product Regulation	Do you agree or disagree with this statement about “light” cigarettes: “Light cigarettes are less harmful than regular cigarettes”?

Table 2.2 Examples of Questions Designed to Measure Policy-Specific Variables in the ITC Surveys
Adapted from Fong *et al.* (2006a)

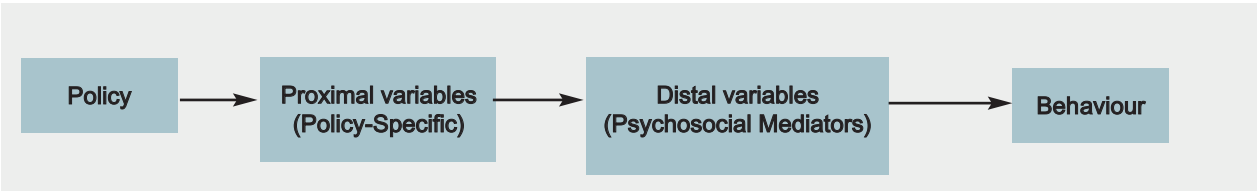


Figure 2.4 Schematic model of how a policy intervention might work (general pathway)

Figure 2.5 Schematic model of how an intervention such as warning labels on cigarettes might work

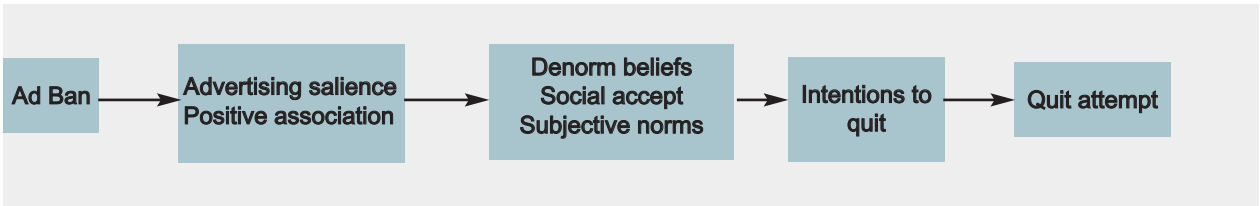


Figure 2.6 Schematic model of how an invervention such as banning of pro-tobacco advertissement might work

mediational model for how it is theorized to operate (Figure 2.4).

For example, an enhancement in warnings may first increase salience/noticing, depth of processing, and other constructs that have been identified by communication theory as being an important initial step for a communication attempt to be effective. The resulting heightened perception of the risk or hazards of smoking should affect overall attitudes and outcome expectancies, which affect intentions, which in turn affect behaviour (Figure 2.5).

In contrast, advertising bans may first decrease awareness of tobacco-favorable messages, which may lead to reductions in the perceptions that smoking is a socially acceptable behaviour, then to the idea that subjective and societal norms are more negative toward smoking, which is theorized to lead to quit attentions and quitting behaviour (Figure 2.6).

The specific articulation of these mediational models leads to specific, theory-driven empirical tests. The strategy of testing the impact of policies through mediational models of this kind differs from the approach taken in dealing with threats to internal validity. That approach, which is a process of *falsification*, uses research design and analytic tools to determine that a possible confounding factor was NOT responsible for the observed pattern of data, whereas explicit

tests of mediational models provide the possibility for confirmatory analyses, which test whether a policy had its impact on an important outcome variable because it first caused changes in a policy-relevant mediator.

In general, the design of the ITC Surveys is guided by the possibility of disentangling the web of alternative explanations and competing forces through the careful selection of specific, theory-driven mediators.

The ITC conceptual model offers an opportunity to test how policies impact or fail to impact anticipated behaviour. For example, the mere existence of a policy, even if implemented properly, does not guarantee that smokers will be exposed to its consequences in the ways anticipated. Using the example of warning labels, some smokers barely look at a pack when they are smoking and may rarely or never notice the warnings. This, however, could be due to motivated avoidance, and it is important to measure whether this has an impact on behaviour. In a cohort survey of Ontario smokers, Hammond and collaborators (2003) found that avoidance of the graphic Canadian warning labels, by means such as covering them up or by putting them in a cigarette case, was not associated at follow-up with a decreased likelihood of a quit attempt.

Additional research questions can be addressed, such as whether is it sufficient for someone merely to notice warnings or whether it is necessary to read them closely, or process them at a deeper cognitive level. And what role do microbehavioural reactions, such as foregoing a cigarette as a result of noticing/reading warning labels, play in determining longer-term outcomes, such as quitting?

In order to address these and other conceptual questions about the impact of warning labels, the ITC Surveys include multiple measures to empirically identify from the service results which measures may be important in understanding the impact of warning labels. In this regard, it should be noted that the ''best'' measure for understanding the impact of warnings may depend on whether the warning is text-based or whether it includes graphic images.

Mediational models have the potential to identify causal mechanisms, and the importance of this is that knowledge of the causal mechanisms can inform the creation of interventions of potentially greater power. Thus, the general mediation model is realized differently in diverse policy domains; different policies are mediated by different constructs. Because the ITC Surveys measure all of these constructs, it is possible to begin to distinguish whether a change of behaviour (e.g. quit attempt) was due to a given policy, in the context of other policies, or to other alternative events that occurred at the same time.

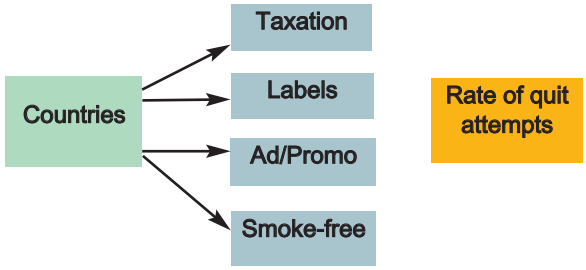
The use of mediational models as a mechanism for establishing the effect of policies:

As described earlier, an important and vexing hazard to internal validity is the concurrent events threat (also known as a *history* threat): the presence of events that occurred concurrently, such as multiple policies, or a mass media campaign that was implemented at the same time as the policy that is being evaluated. How can these threats be measured and dealt with?

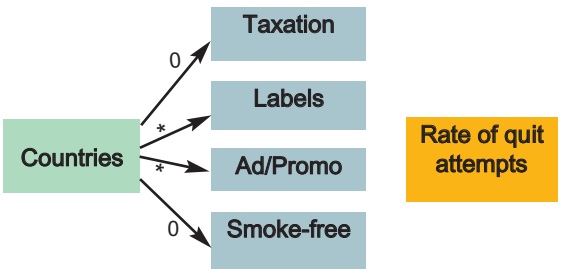
The only method of keeping possible alternative causes from becoming confounders is to measure their potential impact, and explicitly including them in a model that competitively tests their impact. For example, if a mass media campaign is being implemented at the same time as a policy to be evaluated, measures of noticing, and the impact of, that mass media campaign (see Section 5.6) could be included in a post-policy survey, and those measures used as covariates in an analysis of the impact of the policy. Although the study might originally have been conceptualized as evaluating the policy, including measures of the mass media campaign would augment the study as a simultaneous evaluation of the impact of both policy and the campaign. The general point here is that unconfounding of alternative events in the evaluation of a policy can only be attempted through the measurement of the possible impact of those alternative events.

It should also be noted that even randomisation to conditions does not eliminate the threat to internal validity posed by concurrent events. If randomisation were possible in policy evaluation

(a) Basic layout of mediational model designed to test whether any of the policies might have been causally responsible for the difference between countries in the rate of quit attempts.



(b) Between the two ITC survey waves, for each of the four policy domains, did any of the countries make a change?



(c) Between the two ITC survey waves, suppose there were two policy domains in which one country changed: Labels and Ad/Promo (starred paths from countries to those two policy domains). There were no changes over time in the other two domains. Thus, those paths are equal to zero, indicating that differences across countries in the rate of quit attemps could not have been mediated by changes in Taxation and Smoke-free policies.

(d) The reduced mediational model, having eliminated Taxation and Smoke-free policies as possible mediators

(e) We then examine the paths from each of the two policy domains (that is, the policy-specific measures for each of the domains) to rate quit attempts to test whether the change in those policy-specific measures is associated with differences in the Rate of quit attempts. We find that the Label measures are associated with the Rate of quit attemps (indicated by a star), but the Ad/Promo measures are not (indicated by a 0).

(f) Thus, Ad/Promo was not supported as a mediator between countries and rate of quit attempts. That is, changes in Ad/Promo do not help explain why countries varied in quit attempts. In contrast, the significant paths from Countries to Labels and from Labels to Rate of quit attempts supports the contentions that the change in warning labels mediated the pathway from Countries to Rate of quit attemps and that the change in warning labels was responsible for the increase in the rate of quit attemps.

Figure 2.7 The use of mediational models for isolating the effects of specific policies

studies, there would still be the need to measure the impact of other possible influences on behaviour that had occurred between the policy intervention and the post-policy assessment point.

A more complete articulation of the strategy of teasing apart the impact of multiple policies, and/or the presence of other possible influences/confounding factors can be found in the approach to mediational analyses (e.g. Baron & Kenny 1986; MacKinnon *et al.*, 2002; Mathieu & Taylor, 2006; and Spencer *et al.*, 2005). An extended example of the logic of the approach is provided in Figures 2.7 a-f. The scenario is that ITC countries varied in the rate of quit attempts. For simpliciity, four policies are listed: taxation, labels, ad/promo, and smoke free, and the analysis involved the policy-specific variables associated with each of the four policies.

Moderator variables in the ITC Project:

One of the most intriguing lines of inquiry in the ITC Project is to determine whether the impact of the same or similar FCTC policy differs across different countries. In the domain of health warnings (Article 11), the ITC Project is addressing whether the impact of graphic warnings differs across different countries. Among the ITC countries to date, Thailand and Australia have introduced graphic warnings since the beginning of the ITC surveys, and several other countries are anticipated to do so in the future.

The ITC Project is also examining the impact of smoke-free laws in several ITC countries. To date, the impact has been remarkably similar in Ireland (Fong *et al.*, 2006b) and Scotland (Hyland *et al.*, 2007). Ongoing ITC surveys will allow a rigorous comparative evaluation of the impact of smoke-free laws in other ITC countries including France, Germany, The Netherlands and China. Given that the ITC Surveys are using identical or very similar measures and parallel data collection methods across the set of ITC countries, the potential for making conclusions about the commonality or differences of the impact of smoke-free laws, graphic warnings, and the other FCTC policy domains will be strong.

Thus “country” and the environmental and cultural factors that “country” embodies, constitutes an important moderator variable in the ITC conceptual model.

Further, within a country, it is possible to test for differential policy impact on subgroups of a population, by including variables to determine which subgroups are more favourably (and less favourably) influenced by FCTC policies. These moderators fall into five broad classes: *socio-demographics* (age, sex, SES, ethnic background); past behaviour (smoking history, current consumption (cigarettes per day), quit attempts); *personality characteristics* (time perspective, depression, sensation seeking); other environmental effects (stress levels); and *potential exposure to policy* (unemployed people should be less affected by workplace smoking policies).

Dealing with hypothesised moderators is relatively straightforward when they are postulated merely to add predictive power to linear models. The issues become more complex when different mediational pathways are postulated for subpopulations. For example, individuals who avoid warnings might change behaviour through more emotion-related pathways, while those who take in the information on warning labels might be influenced through more cognitive pathways. The ITC Surveys have the design and the measures that will allow the creation of separate models for these different subpopulations, which will make it possible to test whether different subpopulations within a country, as well as between different country populations, respond in the same way or differently to tobacco control policies.

Conclusions

This section has provided some basic principles of how evaluation studies can be designed to offer more confident judgments about the causal impact of tobacco control policies. It has also illustrated the use of study designs (the structural aspects of an

evaluation study) and study features (the selection of measures to be used in an evaluation study, including theoretically guided mediators and moderators).

The eventual outcome of rigorous evaluation studies does not end with a causal statement, however. If mediational analyses demonstrate that a given policy works through changes in one putative mediator but not another, non-policy interventions (e.g. mass media campaigns) can be tailored to influence those mediators that had been identified in the evaluation study to be the operating causal forces leading to favorable changes in behaviour.

Thus, rigorous evaluation of FCTC policies has the potential not only to demonstrate the impact of these policies on tobacco use, but also to provide valuable insights into the development of more effective non-policy efforts to reduce the burden of tobacco use throughout the world.

2.2 Developing and assessing comparable questions in cross-cultural survey research on tobacco

Introduction

The WHO FCTC aims to address the global tobacco epidemic by coordinating national policies to combat tobacco use. This volume illustrates possible conceptual frameworks, methods, and data sets that will be useful for conducting comparative, international research to better understand which policies work and why. This section aims to provide researchers with a basic overview of measurement issues involved in the design and analysis of cross-cultural comparative research, as well as some of the methods currently recommended for attempting to resolve these issues. When possible, we illustrate our points with examples from cross-cultural tobacco research. The organisation of the section follows the general stages of research design, illustrating the corresponding methods used to assess and to avoid introducing systematic measurement error due to cultural differences across the populations in which the research is carried out. The growing literature that we discuss generally reflects concerns related to conducting comparative research across nations and linguistic groups. In most cases, however, the implications and methods we describe extend to intranational studies involving different ethnic groups or even single ethnic groups that speak the same language (e.g. Spanish-speaking Latinos in the USA; people from different socioeconomic groups). In this regard, our general approach may be useful to researchers interested in ensuring the validity of comparative analyses across cultural subgroups within increasingly multi-cultural, intranational settings.

Cross-cultural and cross-national research is often done under the unexamined assumption that question meaning, comprehension and measurement properties are equivalent across cultural groups (Bollen *et al.*, 1993; Smith, 2004a). However, cross-cultural differences in language, social conventions, cognitive abilities and response styles may cause systematic measurement error that biases results in unpredictable ways (Fiske *et al.,* 1998; Harkness *et al.*, 2003a). Apparent differences found across socio-cultural groups may be merely due to measurement artefacts, such as systematic group differences in the meanings ascribed to the same question, whether phrased in the same or different languages. Conversely, true differences may be obscured by such factors as the differential influence of social desirability or the exclusion of items that are important indicators of study constructs in one cultural context but not in another. Whereas the implications of these issues appear most obvious for international comparative research, if left unaddressed, they may also impede our understanding of why certain tobacco policies work better among some socio-cultural groups than among others. In the end, valid cross-cultural comparison demands that measurement error be minimised across the settings and groups of interest (Bollen *et al.,* 1993; Smith, 2004a).

Equivalence of conceptual frameworks

Cross-cultural survey research should begin by assessing whether the conceptual definitions and theoretical frameworks that orient the study reasonably apply across the contexts in which the survey data will be collected. Consideration

of the universal applicability or culturally-specific nature of study concepts is important because their definitions should inform subsequent stages of question selection, development, adaptation and assessment. For example, some concepts may have single or multiple dimensions, each of which should be reflected in its conceptual definition. In some populations the social acceptability of smoking can be characterised by at least two dimensions, one that references close social network members and another that concerns perceptions of a more distal, abstract socio-cultural milieu (Thrasher *et al*., 2006a). These referents may be further subdivided by perceptions of the actual behaviour (i.e. descriptive norms) and desired behaviour (i.e. injunctive or prescriptive norms) (Cialdini, 2003). Hence, at least four dimensions could be delineated within a conceptual definition of the social acceptability of smoking. Nevertheless, the number of dimensions may vary between or within any particular population. Cross-cultural studies should consider construct dimensionality and whether it might differ across cultural groups.

Ensuring the equivalence of concepts across cultural contexts or groups should begin with literature reviews on the topic and concepts of interest. Pertinent literature may nevertheless escape the reach of search engines or the linguistic capabilities of those conducting the reviews, or this literature may simply not exist. This problem may be addressed by establishing collaborative research groups that involve at least one representative from each country or cultural group in which surveys will be conducted (Kuechler, 1987). Ideally, each representative should have native language proficiency and be knowledgeable of both the study topic and the particular contexts in which data collection will take place. Formulating the study's conceptual framework in dialogue among a team of such researchers can help anticipate incongruities in the conceptual framework across survey contexts, and thereby avoid any ethnocentric or universalist tendencies in measurement that might result (Van de Vijver & Hambleton, 1996). Furthermore, this dialogue may help identify cultural or contextual factors that may be important modifiers of tobacco policy effects. Such potential modifiers may otherwise escape consideration because researchers in one context either take them for granted because of their ubiquity or have never considered them because of their absence. For example, strong religious beliefs in some countries may play such a role.

The collaborative process of defining the concepts and framework that orient questionnaire design goes some way toward ensuring that the survey instrument will be meaningful for study participants. There are a number of tensions and difficulties with the collaborative approach, however. As the number of nations or cultural groups involved in the study increases, so do the amount of difficulty and time spent to coordinate efforts and reach consensus (Kuechler, 1987). Granting agencies often demand clearly defined conceptual frameworks before they will fund a project, and without funding to develop this framework, it may be difficult to engage collaborators. The "local" representatives with whom collaboration occurs may actually be quite cosmopolitan, perhaps directly or indirectly socialised into the Western scientific enterprise. Hence, the "cultural" perspective any particular representative provides may be a hybrid form that is at once transnational yet circumscribed by particular social class, gender, and cultural divisions within the country of interest. In this regard, people who have direct knowledge of the local realities of target populations in which survey research will take place may make more substantial contributions toward the development of culturally applicable concepts. Even so, status asymmetries among group members may ultimately overwhelm more local (and perhaps more locally relevant), epistemologies, theories and concepts, particularly if they are incongruent with Western scientific principles (Johnson, 1998). These challenges should be recognised and, to the extent possible, overcome. Collaboration with representatives from each cultural setting nevertheless

forces at least some consideration of cultural particularities and concerns. The resulting conceptual framework should be more likely to "fit" the contexts studied than a framework constructed in the absence of input and involvement of representatives from these different settings.

Question selection and development: equivalence of indicators

The practice of selecting or developing questionnaire items in one language and translating them into other languages is common in cross-cultural survey research. The use of established items saves time, is inexpensive, and allows for ready comparison with other studies that have used the same measures. Ideally, these items will have been pre-tested and found to have suitable measurement properties across subgroups who speak the source language, as well as among those from the linguistic and cultural groups in which the research will be conducted. Such analyses have been done only for a few tobacco survey questions, including those related to dependence (see Section 3.3). If sound measurement properties have been found for the item in one linguistic or cultural context, these properties do not necessarily carry over to the translated version of the item, no matter how good the translation (Harkness *et al*., 2003b). To help ensure equivalence of question comprehension and meaning, pretesting is needed in each major cultural context or major sociocultural group under consideration (see page 68).

One reason why item selection matters is that wording that appears neutral may actually contain phrases or terms with culturally idiosyncratic connotations, making translation difficult (Harkness, 2003). Attempts to capture the meaning of culturally anchored wording—no matter how unambiguous in the original language—may produce awkward translations that violate question design principles and thereby introduce systematic error. One clear example comes from the German General Social Survey item "Das leben en vollen zügen genieβen," which literally translates to English as the nonsensical "Enjoy life in full trains." For American English, a more appropriate translation is the adapted, non-literal phrase "Live life to the fullest" (Harkness, 2003). The often unconscious embedding of cultural anchors in questions may lead to their discovery only through the translation process itself. Similarly, question meanings may not be shared across contexts, and different items will need to be developed in order to adequately reflect study concepts. For these reasons, cross-cultural survey methodologists increasingly argue for methods that open up the translation process to greater scrutiny and more conscious group decision-making (Harkness & Schoua-Glusberg, 1998; Hambleton *et al*., 2005). When cultural anchoring is discovered, unambiguous phrasing in the translated version of the question may necessitate changing the wording of the original language item in order to maintain equivalence (see page 68). Literal question translation may nevertheless result in equivalent meanings across languages. However, it is crucial to consider whether the resulting question adequately captures the concept of interest and whether a non-literal adaptation of the question is necessary to do so (Van de Vijver & Leung, 1997; Van de Vijver, 2004).

Cross-cultural survey research generally involves translating items that are established measures for particular constructs in one language group. For this reason, our next sub-section focuses more intensively on translation approaches. However, researchers may nevertheless consider developing a core set of indicators for use across all sites, supplemented by culture-specific indicators of the same constructs. The selection of culturally-specific indicators should consider measurement research on the same or related concepts conducted within the culture. However, such research may not exist or may involve items that researchers believe are inadequate to capture the meaning of the concept of interest. Item development can follow any of a variety of methods that are standard practice in measurement development, including expert-

driven techniques (DeVellis, 1991) or those that involve eliciting meanings from the target group of interest, as with focus groups (Stewart & Shamdasani, 1998), structured interviews (Spradley, 1979), free-listing, pile sorts and other qualitative techniques (Bernard, 1994; Berkowitz, 2001). Rapid anthropological assessment techniques have also been developed to reduce the time and effort required for more traditional ethnographic methods, with one such effort having already developed a framework for tobacco-related research among youth (Mehl *et al*., 2002). These and other methods could also be used for developing equivalent concept definitions across contexts.

One rarely used approach to item selection and development involves simultaneous, yet independent work by each group responsible for a particular linguistic or cultural subgroup involved in the study (Harkness *et al*., 2003b). This strategy is likely to work best when teams use conceptual definitions that adequately apply across contexts, thereby removing the likelihood that the concepts under consideration are too culturally-specific and, hence, idiosyncratic. Each team would assemble and/or develop items that they believe best reflect the study concepts. In the end, however, incommensurability of items across contexts presents analytic difficulties, as few statistical techniques allow direct comparison of dissimilar stimuli. Furthermore, cross-cultural comparison of only those items with similar content may exclude culturally specific items that are the best and most meaningful indicators of the concept of interest. Overall, this approach involves relatively high development costs, openness to making changes to the source instrument, and complex organisational structure to adequately coordinate teams (Harkness *et al*., 2003b).

Example of focus groups for item development:

Before fielding an international survey of adult smokers in Mexico, in-depth interviews and focus groups were conducted with adult smokers, with discussions oriented by the conceptual domains included in the survey (Thrasher & Bentley, 2006; Thrasher *et al*., 2006a). One concept of interest involved perceived voluntary control over smoking behaviour. This attribution to tobacco consumption behaviour may not only be relevant to self-efficacy regarding quit attempts, but also to perceptions of tobacco products as deviant when compared to other products that people freely decide to consume. When prompted, most all Mexican smokers agreed that tobacco was addictive; however, they found it difficult to explain what "addiction" meant. It became clear that the more common manner of talking about and understanding tobacco's hold over their behaviour was through the term *vicio* or "vice", which connotes a guilty pleasure that is difficult to control, potentially dangerous, and often looked down upon socially. Participants generally agreed that the term addiction, as well as the term *droga* or "drug" also had these connotations. Analyses of data from a subsequent pilot survey of items developed to capture these additional meanings (fumar es un *vicio* ['smoking is a vice"]; *el cigarro es una droga* ["a cigarette is a drug"]) found that these items loaded onto the same dimension as the primary indicator of perceived behavioural control (*tabaco es adictivo* ["tobacco is addictive']), improving the measurement properties of the construct (Thrasher *et al*., 2006a). While the meaning of "a cigarette is a drug" would likely translate back to English, the use of an equivalent English language item that included the term "vice" may be meaningful only within certain subcultural religious groups. As such, this example helps illustrate the development of a culturally-specific item that complements a core item shared across surveys. Cognitive testing of the original item in English and Spanish (see sub-section on Questionnaire Pre-Testing) could complement further statistical analyses (see sub-section on Quantitative assessment) in order to determine whether the single item on vice in the Mexico sample might be used as equivalent to the single item on addiction in samples from other countries.

Approaches to survey translation

Translation of surveys in cross-cultural research is often an afterthought, with little attention paid to the design issues involved in the complex task of producing instruments with comparable measurement properties across languages and contexts (Harkness & Schoua-Glusberg, 1998; Harkness, 2003). Steps described above to ensure the applicability and relevance of construct definitions across diverse contexts provide a foundation for sound translation practices (Harkness *et al*., 2003b). Yet, even with such a framework in place, any of a variety of translation methods could be followed, each with its own advantages and disadvantages. Generally, survey research follows the "Ask-the-Same-Question" model, in which a questionnaire is developed in the "source" language and translated to other "target" languages. Because of its widespread use, we describe methods based on this model, including the "decentering" approach, whose iterative process of translation demands at least some flexibility in the wording of the source language questionnaire.

Ideally, people who translate a questionnaire should be skilled, professional translators who are bilingual in the source and target languages, while having at least some basic training in general principles for developing questions with good measurement properties (for some basic recommendations regarding instrument design, see: Dillman (2007), Bradburn and coworkers (2004) and/or Willis (2005)). If this is not possible, then translation should be conducted by people who are fluent in both languages and practiced in the translation between them. At first glance, a single-person translation appears time- and cost-effective. However, relying on a single person to make all translation decisions may introduce comprehension problems due to regional variance in linguistic expression and meaning, as well as the translator's own idiosyncratic interpretations and inevitable oversights (Harkness *et al*., 2004). Since these issues may result in non-equivalent stimuli and, hence, invalid comparison, the efficacy of single-translator methods increasingly has been called into question (Harkness & Schoua-Glusberg, 1998; Hambleton *et al*., 2005).

A team approach to translation, which involves more than one person who is fluent in the source and target languages, appears to help overcome some biases that result from single-person translations. Team approaches open up to examination and discussion the complex decision-making that occurs in translation, providing a greater range and more balanced critiques of translation options (Guillemin *et al*., 1993; McKay *et al*., 1996; Harkness & Schoua-Glusberg, 1998). Aside from skilled, professional translators (of which there may be more than one), Harkness (2003) suggests that two additional roles be filled in the team approach. *Reviewers* should have language abilities that are as strong as the translators', supplemented with knowledge of questionnaire design principles, study design and the topic of interest. *Adjudicators* should at least share this methodological and topical knowledge, as they will make the final decisions about which translation to adopt, preferably in cooperation with the reviewers and translators who have been more intimately involved in the details of translation and evaluation. When an adjudicator does not understand the source or target language well, Harkness suggests that consultants should be hired to provide this skill. Team approaches involve greater expense, time and coordination than single-person translations; however, this approach is recommended and used by numerous ongoing survey operations, including the Survey of Health Ageing and Retirement in Europe (Börsch-Supan *et al*., 2005), the US Consumer Assessment of Health Care Providers and Systems (Weidmer *et al*., 2006), the US Census Bureau (Pan & de la Puente, 2005) and the European Social Survey (Harkness & Blom, 2006).

The "committee approach" to translation is increasingly viewed as the gold standard in cross-cultural survey research (Harkness & Schoua-Glusberg, 1998; Harkness *et al*., 2004). Generally two to four translators are used, with each additional translator providing more material for critical

discussion of translation possibilities. The *parallel translation* method involves each translator independently translating the same source questionnaire in its entirety. Some of the costs associated with parallel translations can be cut by employing *split translations*, in which each translator is assigned different parts of the source questionnaire. In either case, translators bring their independent translations to a reconciliation meeting where at least one reviewer and perhaps the adjudicator work with the translators to reach agreement on the best translation. The chosen wording could be taken directly from one translation, a mixture of the different phrasings offered, or a previously unconsidered wording that emerges from discussion of the independent translations. Because each question is translated independently by at least two people, parallel translations are likely to offer a greater range of translation possibilities than either split translations or a single translator would produce. The final versions can be adjudicated at the reconciliation meeting or, perhaps provided to the adjudicator for later consideration.

The team approaches to translation may seem extravagant in the context of many low-resource environments. However, the relatively low additional cost of hiring a second translator is likely to offset subsequent costs and data quality issues that might result from an unscrutinised translation. Indeed, this process may anticipate and address questionnaire problems that otherwise only come to light in pre-testing or data analysis. This is not to suggest, however, that this strategy should replace questionnaire pre-testing. Both researchers and translators are likely to come from social strata that differ from the majority of research participants. Hence, translation assessment procedures described below are critical to ensuring sound comprehension and equality of measurement.

Researchers may want to consider allowing for minor changes to the source language questionnaire due to issues that emerge through translation. As described earlier, cultural anchoring of words and phrases may result in translated items that shift original meaning or that violate good question design principles. Either way, systematic measurement error may result. One possible approach to equalising question meaning involves an iterative translation process called "decentering" (Werner & Campbell, 1970). In this method, a source questionnaire provides the starting point for translation to target languages, which could be done using any of the aforementioned methods. However, translators and reviewers signal which items appear to introduce non-equivalence of meaning. Those in charge of each language version of the questionnaire then work in iterative fashion, changing items by tacking back and forth across the translations until all versions appear harmonised. For example, one project using this method translated an English language item that included the term "embarrassed," which existed in the target languages but had stronger connotations than in English. Researchers decided to substitute another term, "unhappy about," which was easier to harmonise across the target languages and did not compromise the measurement properties of the original language item (Eremenco *et al.*, 2005).

The iterative approach to translation is difficult, time-consuming and expensive, and each additional language included in the process will multiply these disadvantages (Harkness *et al.*, 2003b). Unlinking questions from their cultural connotations may result in unwanted ambiguity due to vague, unidiomatic phrasing. Furthermore, changes in source item wording may necessitate pre-testing in order to ensure that measurement properties have not suffered.

Whichever translation approach is taken, we strongly recommend that those involved in cross-cultural tobacco research document their decisions regarding item selection, development and translation. Study concepts should be clearly specified and linked to original, source language items. Translators should be encouraged to keep notes regarding their decision-making processes when translating the item to another language. Similarly, team approaches to translation review should involve further docu-

mentation about how final decisions were made. If the entire questionnaire is not subject to later pre-testing, these notes will help determine which subset of items should be scrutinised more closely. This documentation will also enable future researchers to adequately interpret the data associated with these questions, while providing critical information for further improvement of the measures in later studies.

Example of the committee approach:

One example of the committee approach using parallel translation involves translating an American English-language source survey of adult smokers to the Mexican variety of Spanish. Independent translations of the survey were provided by four bilingual professional translators, three of whom were Mexican nationals and the fourth an American who had been living in Mexico for 19 years and working as a professional translator for 24 years. Although all of them had at least some experience with survey translation, each was provided with summary materials on question design principles and asked to follow them. Two of the Mexican translators were recruited because they were regular smokers, as was a young adult, bilingual Mexican research assistant who had been involved in earlier stages of the project and who served as a reviewer at the reconciliation meeting. As members of the target population in which the survey would be administered, these three people helped ensure the use of natural terminology and comprehensibility among smokers. Because of logistical and cost constraints, representatives were not included from each of the different regions of Mexico where the survey would be administered. This was a potential limitation.

The reconciliation meeting involved a full day of work with three translators (one was unable to make the meeting but provided her independent translation), two bilingual reviewers, and a bilingual reviewer/adjudicator. After beginning the session with a further discussion of question design principles, we examined the original English version and all four translations, addressing one question at a time. As emphasised in the description of the methodology, this process produced a range of possible translations, even for questions that, on the surface, appeared straightforward. The beginning of the process was time-consuming and challenging. However, decision-making became easier as participants became comfortable with the process and as we reached agreement on terms, grammatical structure, and response options that were repeated throughout the questionnaire.

As an illustration of the decision-making processes involved in this method, the following describes how we translated the last phrase of the question “On average, how many cigarettes do you smoke each day, including both factory-made and roll-your-own cigarettes?” This clarification to this standard question had been included in the source language questionnaire in order to ensure that respondents considered “roll-your-own” cigarettes, particularly as switching to lower-cost tobacco is a common response to raising the price of cigarettes (Young *et al*., 2006).

One non-smoking translator deleted the last clause of the English version because she had never heard of people using such cigarettes in Mexico. However, we did not want to exclude mention of this practice since it occurs in Mexico, although at a low prevalence. Indeed, one aim of the survey was to estimate this prevalence, although it would be measured with more precision in a question that appeared later in the survey instrument. Two general options for describing factory-made cigarettes emerged: one was a more literal translation (*cigarros hechos en fábricas*, literally “cigarettes made in factories”) and the other turned the focus toward branded and marketed cigarettes (*cigarros de marcas comerciales*, literally, “commercial cigarette brands”). This second focus was discarded since rolling tobacco is also branded and marketed, even though unbranded, loose tobacco can be bought in some regions of Mexico. The more literal translation sounded awkward and seemed to divert attention from the main question content. In the end, we decided on a phrase that could be roughly translated as “cigarettes from the pack”

(*cigarros de cajetilla*), since the word for pack (*cajetilla*) connoted "factory-made" without sounding awkward, while setting up the contrast with the "roll-your-own" type cigarettes that would be mentioned thereafter.

For the final clause in the question, two options emerged from the three independent translations. One used a term for rolling that is also common for rolling marijuana cigarettes (*cigarros forjados a mano*) while the other introduced the participant as the one who "made" (hacer) the cigarettes (*cigarros hechos por* usted, literally "cigarettes made by you"). There was agreement that either option could confuse people who did not engage in rolling cigarettes — this would be the vast majority of study participants. However, reference to the participant making the cigarettes seemed on track, since not including the participant as agent could cause people to think of cigars, which are also hand rolled, but by someone else. We agreed on a longer version "cigarettes that you make by hand" (*cigarros que usted hace a mano*). Later cognitive interviews indicated that this phrase nevertheless connoted marijuana cigarettes for some participants, and so the final, pretested version clarified that these were cigarettes made with tobacco: En general, *¿cuántos cigarros al día fuma, incluyendo los cigarros de cajetilla y los cigarros de tabaco que usted hace a mano?* (Literally, "In general, how many cigarettes do you smoke each day, including cigarettes from the pack and tobacco cigarettes that you make by hand?"). Finally, interviewer training included a focus on the meaning of the question, so that interviewers could anticipate and respond to any comprehension difficulties that they sensed among participants.

This example illustrates a number of the advantages that accompany the committee approach to translation. Importantly, there were a variety of options to choose from. Consistency of terminology and phrasing across translation options would have provided support for selecting a particular translation. The example above indicated inconsistencies in the terms and wording, which led to group decision-making about the best way to resolve discrepancies. Moreover, resolutions to discrepancies did not appear in the originally translated versions. Finally, the version agreed upon in the reconciliation meeting still needed to be altered a little after cognitive testing indicated undesirable connotations for one part of the question.

Culturally moderated response styles

Comparisons across cultural groups may be biased by systematic differences in "response styles," such as social desirability, extreme responding, and acquiescence. Of particular concern are social desirability effects, which manifest when respondents misrepresent or edit their true responses to a question in order to project an image of themselves that accords with their perceptions of social norms and expectations (Marlow & Crowne, 1960). The phenomenon appears to be universal across societies, with stronger effects found when considering self-report of behaviours or beliefs that are socially sanctioned within a given cultural context (Johnson & Van de Vijver, 2004). Hence, the differential effects of social desirability on self-reported tobacco attitudes, beliefs, and behaviours should be proportional to the level of tobacco's social unacceptability across the socio-cultural groups under consideration. Because social desirability effects also appear stronger among minority or disenfranchised groups within a society (Ross & Mirowsky, 1984; Edwards & Riordan, 1994; Warnecke *et al.*, 1997), it may disproportionately influence national samples that contain more minority group participants.

Social desirability appears positively correlated with a number of macro-level societal characteristics, such as higher levels of "collectivism" and lower levels of "individualism." Higher levels of social desirability appear congruent with, and may stem from, collectivist codes of social interaction that emphasise courtesy, maintaining harmonious relations and saving face (Marín & VanOss Marín, 1991; Johnson & Van de Vijver, 2004). Smokers from collectivist societies that stigmatise tobacco use may view true representation of their thoughts and behaviours in an

interview context as threatening these more important elements of social interaction. On the other hand, people from individualist societies appear to have stronger prohibitions against providing misleading information (Triandis, 1995). Hence, smokers in these societies may be less likely to provide socially desirable responses independent of the extent of social sanctions against smoking. This suggests that individualism/collectivism and social sanctions against tobacco are likely to interact, producing differential social desirability effects on tobacco survey questions. The strongest effects of social desirability should occur under conditions of strong stigmatization of smoking behaviour in a collectivist society, whereas the weakest effects would occur in individualist societies with weak stigmatisation. Future research should empirically test this proposition.

Several other response styles have also been found to vary across cultures (Baumgartner & Steenkamp, 2001). Two that have perhaps received the most attention are extreme response styles (Smith, 2004b) and acquiescence (Knowles & Condon, 1999). Extreme response styles refer to the greater preference of respondents from some cultures to select the most extreme endpoints of response scales, whereas respondents from other cultures are more likely to make less extreme choices when answering. Moreover, some respondents exhibit a greater tendency to agree with questions read by interviewers, even when the questions are contradictory, a process referred to as acquiescent responding.

Although there is general agreement that social desirability, extreme responding and acquiescence are each moderated by culture, there is less consensus or available evidence regarding how to best account for these potential sources of measurement error when conducting cross-cultural research. Several researchers have attempted to neutralise social desirability effects by explicitly measuring these propensities and then statistically adjusting for them (Nederhof, 1985). Most reported attempts to introduce social desirability corrections, however, have been unsuccessful (Ones *et al.*,1996; Ellingson *et al.*, 1999; Fisher & Katz, 2000), suggesting that other approaches should be explored (for reviews of other methods of addressing social desirability in survey research, see Nederhof (1985) and Paulhus (1990)). Some researchers have also reported studies in which they assessed extreme responding and/or acquiescence via structural equation modelling (Mirowsky & Ross, 1991; Greenleaf, 1992; Watson, 1992; Billiet & McClendon, 2000; Cheung & Rensvold, 2000). In general, however, there is no consensus on how to best confront problems of systematic cross-cultural variability in survey response styles.

During data collection, efforts are also commonly made to minimise the social distance between respondents and interviewers by attempting to match them on ethnic background or demographic characteristics in hopes of minimising the social desirability pressures placed on respondents. For example, in contexts where deference to authority is a key cultural value, interviews conducted by older people of higher social status may induce strong social desirability effects. Numerous studies are available that demonstrate respondent deference to interviewers who represent differing cultural backgrounds (Cotter *et al.*, 1982; Anderson *et al.*, 1988; Finkel *et al.*, 1991; Davis, 1997; Johnson *et al.*, 2000), although it should be noted that none of these studies are based on experimental evidence. Under some circumstances, too little social distance between respondents and the person interviewing them may encourage socially desirable responding (Dohrenwend *et al.*, 1968). Concern with the effects of social distance can also be extended to interview mode, as the degree of privacy afforded by each mode of data collection may exert differential pressures on respondents to provide socially desirable information. Although little information is available with which to examine cultural variability in mode of interview effects (Marín & Marín, 1989), it would seem likely that the social sensitivity of the answers being requested and respondent culture might interact with survey mode in ways that either magnify or

minimise substantive differences across groups. These effects may be difficult to predict, particularly given the near absence of research on this topic. Researchers should thus carefully consider how the social sensitivity of the topics examined might vary across the groups studied, the types of questions asked, and how the mode of data collection might influence participants' responses.

Questionnaire pre-testing and translation assessment

We focus on two approaches to questionnaire pre-testing and translation assessment. First, we discuss back-translation, which has been used frequently and even viewed as a gold standard for translation assessment; however, we describe a number of pitfalls that recommend against its use as a sole assessment method. Second, cognitive interviewing is described, since it is increasingly recognized as a crucial pre-testing stage before surveys go into the field within particular socio-cultural settings. We suggest that the rationale in favour of this approach be extended to support the use of cognitive interviewing to assess translated questionnaires. Another method for determining comprehension and meaning attributed to items involves focus group evaluation with members of the target population. This assessment approach is likely to be better than no pre-testing of the survey instrument; however, the information from cognitive interviews may be of higher quality because it better approximates the dyadic interplay of survey administration than do focus group dynamics. Finally, another promising tool for assessing respondent cognitions related to translated questions is behavioural coding, a technique which codes respondent and/or interviewer reactions to questions in recorded interviews to identify problematic survey questions (Fowler, 1995; Van der Zouwen & Smit, 2004; Johnson *et al.*, 2006). Overall, we emphasise the importance of translation assessment and pre-testing as a means of ensuring sound measurement properties of the target language survey instruments.

Back-translation:

Back-translation is often mistaken as a method of translation, but it is actually a method for assessing the quality of a translation that has already been made into a target language (Harkness, 2003). It involves independent translation of the target language questionnaire back into the source language and comparing the result with the original source language questionnaire. Back-translation presumes that the greater the similarity between the results, the more acceptable the translation (Brislin, 1970). However, languages are not isomorphic, and an unnatural sounding or even incomprehensible target language translation may produce, or even be necessary for, a "good" back-translation. Although back- translation may reveal some problems with target translations, it does not adequately assess the translated questions' comprehensibility within the target population (Harkness & Schoua-Glusberg, 1998; Harkness, 2003). Furthermore, the methodology provides no guidance about what qualifies as an acceptable level of similarity across the source and back-translated versions. Finally, when a back-translated questionnaire depends on a single translator for the "forward" translation into the target language—as it often does—it neither opens up the translation process to critical scrutiny nor does it produce the range of translation options that are found in team approaches. These factors recommend against the use of back-translation as the only method of translation assessment. Translation quality also needs to be evaluated in a more direct fashion.

An example provided earlier helps illustrate these concerns. The German General Social Survey item "Das leben en vollen zügen genieβen" literally translates to English as "Enjoy life in full trains." This translation is readily back-translated to and reproduces with fidelity the original German source language phrase. However, the nonsensical nature of the English translation could go undetected without further review. Moreover, an appropriate British adaptation of this phrase ("Live life to the full") would sound awkward in American English, for which different wording would be necessary (i.e. "Live life to the

fullest."). Such nuances would be missed, and in fact be discouraged, with back-translation that did not entail further review by bilinguals (Harkness, 2003).

Cognitive interviewing:

Cognitive interviewing is increasingly used to pre-test and thereby improve comprehension and related measurement properties of questionnaires within particular societies (Willis, 2005). The rationale for and principles that orient this practice should extend to assessment of translated questionnaires. In the absence of such pre-testing, there is no guarantee that the target language instrument will have sound measurement properties, even when the instrument has been pre-tested in the source language and best practices have been followed when translating it (Harkness *et al.*, 2003b). We describe a few basic principles of cognitive interviewing, while referencing key works for readers who are interested in more detail.

Cognitive interviewing follows from research on the cognitive processes involved in responding to survey questions (Willis, 2005). The response process generally involves question comprehension (i.e. meaning of terms and perceived intent of question), retrieval from memory (i.e. availability of and strategies to access relevant information), judgment processes (i.e. motivation to respond and to respond truthfully) and mapping the internally generated response to the question onto the response categories provided. As each step along this pathway may introduce measurement error, cognitive interview techniques focus on these aspects of the recall process.

The "think aloud" and "verbal report" protocols generally involve asking participants to openly describe the stream of thought in which they engage as they answer a survey question (Ericsson & Simon, 1984; Conrad & Blair, 2004). Responses are usually audio-recorded and transcribed for analysis. Advantages of the method include the minimal training requirements for the interviewer, whose main task is simply to read the question and listen. This generally passive interviewer stance may result in lesser bias than more pro-active methods. However, although the open-ended format of this approach may allow unanticipated response issues to emerge, subjects may need to be trained to think aloud, with some people unable to develop the skills necessary to provide useful feedback. Even "good" participants wander off track, thinking in ways that may only vaguely correspond with the mental processes required to respond to the question under normal circumstances (Willis, 2005).

Verbal probing techniques are increasingly favoured over think-aloud strategies in cognitive interviews (Willis, 2004, 2005). Probes have been developed in accordance with principles of sound question design, with specific probes used to uncover specific processing issues (see Table 2.3). An interview protocol is generally developed to anticipate which kinds of probes, if any, will be necessary for each question. However, the interviewer may also freely employ probes to address issues that unexpectedly emerge during the course of an interview. As such, the use of verbal probes demands the active involvement and training of the interviewer. However, training is less of an issue for the survey respondent than in the think-aloud. Probes may nevertheless influence respondents in ways that do not adequately reflect cognitive processes under "real" survey conditions. In particular, care must be taken to develop unbiased, neutral probes that do not lead participants to respond in particular ways.

When addressing survey instruments within particular socio-cultural settings, Willis (2005) recommends that each round of cognitive interviews involve survey administration among 8 to 12 people from the target population. At least two testing rounds are necessary to assess the adequacy of the original questionnaire as well as changes that result from the first round. Although the number of testing rounds will depend on the quality of the original instrument and the proposed revisions, Willis suggests that there are likely to be diminishing returns after three rounds of testing. This may or may not be the case in dealing with more complicated cross-cultural issues that involve translated questionnaires, where each round

READING: Is it difficult for interviewers to read the question in the same way to all respondents?
- What to read: interviewer may have difficulty determining what parts of the question to read
- Missing information: information that the interviewer needs to administer the question is not provided
- How to read: question is not fully scripted and therefore difficult to understand

INSTRUCTIONS: Look for problems with any introductions, instructions or explanations from the respondents' point of view
- Conflicting or inaccurate instructions, introductions or explanations
- Complicated instructions, introductions or explanations

CLARITY: Identify problems with communicating question intent or meaning to the respondent
- Wording: question is lengthy, awkward, ungrammatical or contains complicated syntax
- Technical terms: terms undefined, unclear or complex
- Vague: multiple ways to interpret the question or to decide what is to be included or excluded
- Reference periods: missing, not well specified, or in conflict

ASSUMPTIONS: Determine problems with the assumptions made or underlying logic
- Inappropriate assumptions are made about the respondent or about his/her living situation
- Assumes constant behaviour or experience for situations that vary
- Double-barrelled: contains more than one implicit question

KNOWLEDGE/MEMORY: Check whether respondents are likely to or not know or have trouble remembering information
- Knowledge may not exist: respondent is unlikely to know the answer to a factual question
- Attitude may not exist: respondent is unlikely to have formed an attitude about the argument being asked about
- Recall failure: respondent may not remember the information asked for
- Computation problem: the question requires a difficult mental calculation

SENSITIVITY/BIAS: Assess questions for sensitive nature or wording and for bias
- Sensitive content (general): the question asks about a topic that is embarrassing, very private, or that involves illegal behaviour
- Sensitive wording (specific): given that the general topic is sensitive, the wording should be improved to minimize sensitivity
- Socially acceptable: a socially desirable response is implied by the question

RESPONSE CATEGORIES: Assess the adequacy of the range of options
- Open-ended question: is inappropriate or difficult to answer without categories to guide
- Mismatch: question does not match response categories
- Technical terms: are undefined, unclear or complex
- Vague: responses categories are subject to multiple interpretations
- Overlapping: categories are not mutually exclusive
- Missing: some eligible responses are not included
- Illogical order: order not intuitive

ORDERING OR CONTEXT problems across questions

Table 2.3 Questionnaire Design Issues, from Willis (2005)

Adapted from Willis & Lessler (1999) and Willis (2005)

would be followed by efforts to coordinate and translate questionnaire changes until any cross-group discrepancies in question interpretation and comprehension appear to be resolved.

Where equivalence of meaning cannot be achieved, researchers should document why, and make sure this documentation is accessible to those who will ultimately analyse the data. Researchers who use the data at a later date may otherwise believe that the questions are equivalent and make invalid comparisons across cultural groups. Drawing from the previous example regarding the "vice" connotation of "addiction" in Mexico (see page 62), it may be inappropriate to compare Mexican smokers' and smokers from other countries on the item "tobacco is addictive" if the dominant meaning of addiction is compulsive behaviour in other countries. This situation could be documented by describing how "addiction" in Mexico appears to more strongly connote vice and less strongly denote compulsion than in other countries.

Cognitive interviewing example:

One recent example of cognitive interviewing to pre-test translated items involved the Spanish version of the Adult Tobacco Survey (ATS) for the United States' National Center for Health Statistics and the Office on Smoking and Health at the Centers for Disease Control and Prevention. The goal was to produce a Spanish-language version of the ATS questionnaire that was equally comprehensible and that shared the same meaning among Latinos in the US who speak different national varieties or dialects of Spanish. In the first step, a committee approach was used involving independent, parallel translations by bilingual translators of Mexican, Puerto Rican and South American heritage. This was followed by two rounds of cognitive interviews with Latinos from nine countries and Puerto Rico. The first round involved 40 participants using "think-alouds" after every question. In the second round, the resulting survey was administered in normal fashion to 28 participants, followed by a debriefing that targeted particular comprehension issues.

One of the many issues that came up concerned the translation of the often-asked English-language question, "Have you smoked 100 or more cigarettes in your life." Participants repeatedly thought that this question referred to daily smoking, even after the word "entire" was inserted to read "in your entire life" (*en toda su vida*) and the phrase was printed in boldface type to ensure its emphasis by survey administrators. This underscores the point that modification of a question may not resolve the problem, hence modified versions should also be pre-tested (Forsyth *et al.*, 2004). To resolve the issue, an introductory phrase was added to both the English and Spanish language questions: "For this question, we want you to think of all the cigarettes you ever smoked in your whole life, not on a single day." In this case, changes made to the Spanish-language items meant re-evaluating and changing the wording of the original, English-language version in order to reinforce equivalence. Anecdotal evidence suggests that similar comprehension problems characterised the original English-language version, so the addition of this introductory phrase may have improved comprehension across languages.

Quantitative assessment of measurement properties and systematic measurement error

Despite all precautions to ensure item equivalence across social-cultural groups and linguistic variants of a questionnaire, some unaccounted-for factor may nonetheless systematically and differentially influence responses provided by the groups under consideration. The strategies described here are best employed after collecting pilot data, but before implementing the full survey. Results can be used to eliminate, change or replace items that appear to be biased. However, these methods can also be used to assess measurement equivalence after survey data are collected, with the drawback that it is too late to change items with poor measurement qualities. As has been

emphasised when addressing other measurement equivalence issues described in this section, it is recommended that such issues be documented so that others who use the data at a later date will be aware of these issues.

Three approaches are briefly described here: single indicators, "alternative indicators" and latent variable Structural Equation Modeling (SEM). When multiple indicators of a construct are used, more statistical means are available to try to rule out systematic measurement error across groups. However, some approaches demand that single constructs be measured with a large number of items, which makes them less applicable to survey research. These methods, such as multi-trait multi-method (Saris, 2003a), multi-dimensional scaling (Fonatine, 2003), and item response theory approaches (Saris, 2003b) are detailed elsewhere.

Single-item measures of constructs:

When a single item is used to measure a construct, it may be difficult to assess whether observed similarities or differences in the measure are valid or whether these observations result from some other nuisance factor. Differential patterns of item non-response or "do not know" may indicate non-equivalence. Indeed, these non-random patterns violate assumptions that are necessary when dealing with this issue through pairwise or listwise deletion, as well as when using multiple imputation techniques (Groves, 2001). Nevertheless, theory and previous empirical findings can be drawn upon in order to predict how the indicator should correlate with other variables. In other words, expected correlations with other particular variables provide evidence of convergent validity. The absence of such correlations does not necessarily disprove the validity of the measure, however. Rather than disconfirming the validity of the measure, this lack of correlation may instead merely indicate the inadequacy or general inapplicability of the theory. Indeed, even when the measure under consideration is correlated with a set of theoretically related variables, this merely provides evidence — not confirmation—of the measure's convergent validity; systematic measurement error across the theoretical set of variables may still bias group comparisons.

Alternative measures of the same construct:

When there are multiple indicators of a particular construct, differential item functioning across cultural groups can be assessed by alternatively considering each indicator (Bollen *et al*., 1993; Smith, 2004a). With two items, a relatively clear indication involves consistent results for group differences in means (e.g. both higher in one group versus another) and in correlations with other constructs (e.g. number of days and number of cigarettes per day correlated with addiction). If the two indicators show inconsistent results, then strong claims about either result will depend on one's ability to convincingly argue for the use of one indicator over another. Although such post-hoc argumentation may be suspect, it can also establish the focus for subsequent research to clarify measurement and the interpretations that result. With three alternative indicators of the same construct, results from the third indicator can tip the balance in favour of the "preponderance of evidence." Consistency across all three indicators provides relatively strong confirmation of the validity of the results. Smith suggests that the most robust evidence will come from consistent results across alternative indicators that not only contain linguistically different stimuli, but that also have different response formats (Smith, 2004a).

Simultaneous assessment of multiple indicators:

Data collection on multiple indicators of the same construct also allows for statistical assessment of all indicators simultaneously, instead of the sequential format outlined above. Simultaneous consideration of multiple indicators lessens the impact of idiosyncratic, and therefore problematic, indicators (Bollen, 1989; Bollen *et al*., 1993). It also allows for the application of more formal statistical procedures to test, improve and attempt to equalise construct measurement properties across groups.

Exploratory factor analysis (EFA) techniques can provide evidence for the equivalence of construct dimensionality and discrimination across groups, although special techniques are often necessary to ensure adequate comparison (Van de Vijver & Leung, 1997). Items may be considered for elimination if substantial group differences are found for factor loading values on the same dimension or for the extent of cross-loading across dimensions. Cronbach's alpha may also be used to determine group differences in inter-item reliability. Although some statistics are available for evaluating factorial agreement across groups, the sampling distributions for these statistics are unknown, hence there are no statistical means of testing for what counts as an unacceptable difference (Van de Vijver, 2003). Moreover, these techniques generally assume normally distributed, continuous variables, and survey indicators often violate these assumptions.

Latent variable structural equation modelling (SEM) offers a more direct means of testing invariance of construct parameters and measurement properties across groups (Bollen, 1989, 2002; Joreskog & Sorbom, 1996). As with EFA, the dimensionality of different concepts can be examined. However, a key advantage of SEM concerns the ability to use statistical tests of construct parameter equivalence across groups. Moreover, whereas factor analysis parameter estimates assume continuous, normally distributed indicators, SEM allows estimation using non-normally distributed categorical and ordinal indicators (Joreskog & Sorbom, 1996; Muthen & Muthen, 2004). SEM techniques estimate items' unique weighted contributions toward the measurement of latent variables. EFA, on the other hand, involves summing or averaging variables that comprise a particular dimension, treating each indicator as equally weighted. Finally, several SEM packages now adjust for study design effects and sampling weights—adjustments that are often important in generating reliable, unbiased estimates in cross-cultural survey research. Taken as a whole, these key advantages recommend SEM methods over standard EFA techniques. Cepeda-Benito and colleagues (Cepeda-Benito *et al.*, 2004) provide a recent example of the use of these models to compare the structure of the Questionnaire of Smoking Urges survey instrument across samples of American and Spanish smokers.

Summary and Recommendations

Evaluation of tobacco control policies and other population-level interventions often involves data collection efforts across diverse national, cultural, linguistic and social groups. Comparison across such groups is often necessary to clarify policy effects, how these effects happen, and how effects might differ across populations. The literature discussed in this section suggests that these comparative studies should consider measurement equivalence issues in the following ways:

- Research teams should include collaborators from the socio-cultural groups in which the study is being conducted in order to help anticipate issues regarding the comparability of the theoretical framework, constructs and the measurement of these constructs across groups. When research involves participants from distinct language groups, at least one, and preferably more, team members should be fluent in the source language and the target language in which the survey will be administered.
- Whenever possible, it is recommended to use measures that have been appropriately validated for the populations in which the questionnaire will be administered. Even when a measure has been validated within one population group, its validity may not extend to other groups, and additional steps may be necessary to increase validity and improve the value of comparisons across groups.
- Translation of questionnaire items from one language to another should involve experienced translators. Review and adjudication of multiple, independent translations of the same items is currently considered the gold standard. If only one person translates

the questionnaire, then translation review should involve a group of bilingual people who are knowledgeable of questionnaire design principles and of key study concepts. Translation assessment should not merely consist of backtranslation.

- Researchers should carefully select and translate items with the goal of achieving equivalence of construct meaning across study populations. In some cases, literal translation of a questionnaire item across linguistic variants of the survey will not adequately capture the construct of interest, and more flexible translation and adaptation of the question will be necessary.
- All surveys, not just those that are translated, should be pretested to assess comprehension issues among the populations in which the survey will be administered. Ideally, pre-testing would involve cognitive interviewing before a survey is fielded. Cognitive interviewing or other pre-testing methods may also be used post-hoc to increase the validity of comparisons or to determine whether inconsistent results may be due to differential question comprehension.

  Researchers should consider and seek solutions to minimise the ways in which culturally moderated response factors (e.g. social desirability, acquiescence, extreme responding) may influence responses.

  Researchers should document decisions related to measurement development and item wording, especially where conceptual equivalence is suspect, translation is difficult, or where cognitive interviewing or other pre-testing methods reveal systematic differences in meaning. Researchers should also document issues around survey administration.

3.1 Measuring tobacco use behaviours

Introduction

The majority of tobacco control policies are designed to reduce tobacco use or exposure to tobacco smoke in the environment; strategies that are clearly supported by the scientific literature (US Department of Health and Human Services, 2004, 2006; IARC, 2004, 2007a). Preventing initiation and promoting quitting are the two major tobacco control strategies designed to reduce use. To facilitate progress, article 20 of the WHO Framework Convention on Tobacco Control (FCTC) calls for Parties to:

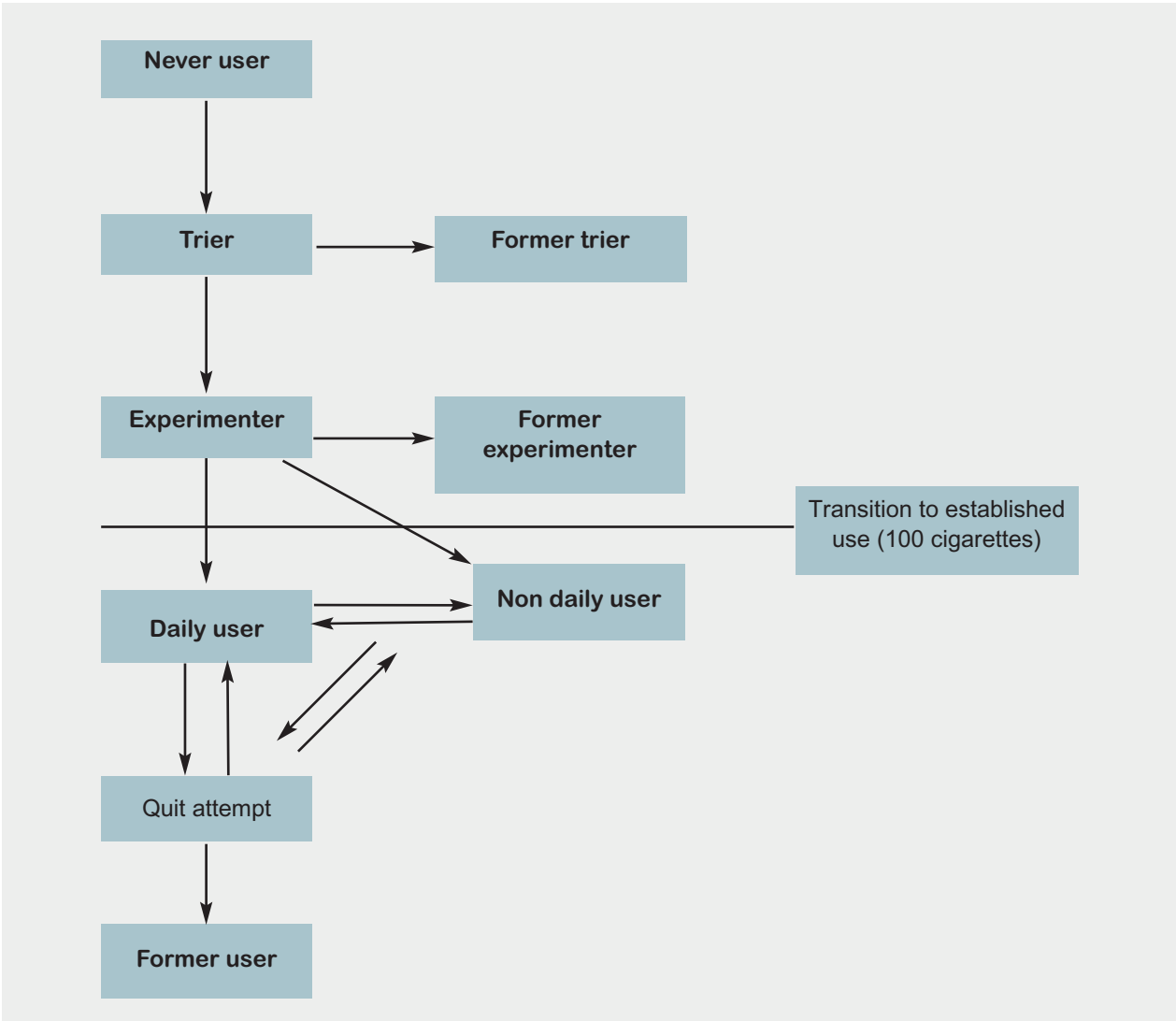
"(a) establish progressively a national system for the epidemiological surveillance of tobacco consumption and related social, economic and health indicators
(b) cooperate with competent international and regional intergovernmental organizations and other bodies, including governmental and nongovernmental agencies, in regional and global tobacco surveillance and exchange of information on the indicators specified in paragraph 3(a) of this Article
(c) cooperate with the World Health Organization in the development of general guidelines or procedures for defining the collection, analysis and dissemination of tobacco-related surveillance data."

In addition, Section 1-d of Article 21 requires each ratifying nation to provide periodic updates on surveillance and research as specified in Article 20. Article 22 calls for cooperation among the Parties to promote the transfer of technical and scientific expertise on surveillance and evaluation, among other topics (WHO, 2003).

This section will first review the natural history of tobacco use (e.g. initiation, current use, cessation). In epidemiologic studies of disease etiology, such as those discussed in IARC Monographs (e.g. IARC 2004) and reports of the Surgeon General (US Department of Health and Human Services, 2004), tobacco use behaviours (e.g. number of years smoked, number of cigarettes consumed each day) serve as independent variables. In the evaluation of the tobacco policies discussed in this Handbook, tobacco use behaviours serve as dependent variables. The section will then discuss factors that can influence the validity of self-report and factors that can influence comparability across surveys. The section will end by describing several measures to assess use, providing examples from cross-national surveillance and evaluation systems (Section 4.3), as well as national sources.

Natural history of tobacco use

The natural history of tobacco use is often conceptualized as a series of steps that can progress from never use, to trial, experimentation, established use, attempting to quit, relapse, and/or maintenance of cessation (Figure 3.1 and Table 3.1) (US Department of Health and Human Services, 1990, 1994; Marcus *et al.*, 1993; Pierce *et al.*, 1998b; Mayhew *et al.,* 2000; Choi *et al.*, 2001; Hughes *et al.*, 2003). Prior to actual initiation of use, never users often think about use, a step in the process that is described in Section 3.2. After initial trial, users can either continue to experiment or discontinue and become former triers. Experimenters can either progress to established user or discontinue use and become former experimenters. Recent research suggests that nicotine dependence may appear during the experimentation phase, before use becomes established (DiFranza *et al.,* 2002a; O'Loughlin *et al.,* 2003; Fidler *et al.,* 2006). Use becomes established when a threshold of cumulative lifetime exposure is surpassed. The exact threshold of established use is unknown and likely varies considerably, but is often considered as having smoked at least 100 lifetime cigarettes, or being exposed to a similar amount



Note: "Use" involves consumption of cigarettes, other forms of smoked tobacco products, and/or various smokeless tobacco products.

Figure 3.1 The natural history of tobacco use

I. **Initiation**
- a. Intention to try (Section 3.2)
- b. Initial trial
  - i. Discontinuation after initial trial
- c. Experimentation
  - i. Discontinuation of experimentation

II. **Transition to established use**
- a. Ever daily versus never-daily

III. **Current use**
- a. Frequency of use (daily versus non-daily)
- b. Type of product used
- c. Brand used
- d. Intensity of use (units/day)
- e. Topography (for smoked products)
- f. Purchase patterns (partly covered in Section 5.1)

IV. **Cessation**
- a. Intention to quit (Section 3.2)
- b. Quit attempt
  - i. Intentionality
    1. Planned
    2. Spontaneous
  - ii. Dose management
    1. Abrupt discontinuance
    2. Gradual reduction
  - iii. Methods (Section 5.7)
    1. Assisted
    2. Unassisted
- c. Maintenance of abstinence versus return to use

†Here the term "use" means consumption of cigarettes, other forms of smoked tobacco products, and/or various forms of smokeless tobacco

Table 3.1 The Natural History of Tobacco Use†: Key Constructs

of other tobacco products. Established use is generally manifested as daily use. However, persistent, regular non-daily use can also take place (Evans *et al.,*1992; Husten *et al.,* 1998; Trosclair *et al.,* 2005). Once past the threshold of established use, discontinuance involves an attempt to quit, with the outcome of each quit attempt being either relapse or maintenance of cessation (US Department of Health and Human Services, 1990; Gilpin & Pierce, 1994; Hughes *et al.,* 2003; West, 2006). Quit attempts can be planned or spontaneous, involve abrupt discontinuance or gradual reduction in use before quitting, and may or may not be assisted by one or more of several available treatment strategies (Fiore *et al.*, 1990; Giovino *et al.,* 1993; West, *et al.*, 2001).

Validity of self-report of current tobacco use behaviours

Survey-based measures of current tobacco use behaviours, assessed in samples that are representative of a given population, allow researchers and policy-makers to estimate patterns of and trends in use overall and for subgroups in the population. National prevalence estimates have, in the vast majority of cases,

been based on self-reports of personal behaviours. Self-report, however, may be subject to misclassification bias. Survey respondents can either state that they do not currently use tobacco, when in fact they do (misclassification of use as non-use), or that they do currently use tobacco when, in fact they do not (misclassification of non-use as use). Each of these misclassification biases can compromise the validity of a survey estimate.

Determining validity:

Validation of self-report is generally conducted using biomarkers of exposure to tobacco or tobacco smoke as criteria. Biomarkers of exposure that have been used in studies include nicotine; cotinine, a major metabolite of nicotine; carbon monoxide; and thiocyanate (Society for Research on Nicotine and Tobacco, 2002; Al-Delaimy, 2002). Nicotine and cotinine are almost exclusively specific to tobacco products. Very low levels of nicotine can be found in some vegetables, but their impact on cotinine levels is insignificant (Pirkle *et al.,* 1996; Society for Research on Nicotine and Tobacco, 2002). Cotinine is preferred over nicotine as a biomarker, because it has a longer half-life in biological fluids than nicotine (~16 hours versus ~2 hours), thus reflecting use over the previous three days for the general population (Society for Research on Nicotine and Tobacco, 2002). Cotinine can be obtained from saliva, urine, and blood (serum). Saliva is the biological fluid of choice in population-based surveys, because it is the easiest to obtain. Hair nicotine levels reflect exposure over a longer period of time (Al-Delaimy, 2002). Hair samples are even easier to obtain than saliva. However, measurement of nicotine in hair can be influenced by hair color, treatment, and growth rate and identifying nicotine from actual tobacco use versus exposure to environmental sources can be problematic (Al-Delaimy, 2002).

Unfortunately, the use of biomarkers as indicators of actual use is also subject to error. Studies using cotinine to validate self-report must determine a cut-off for discriminating users from non-users. Cut-offs generally range from 10.0-20.0 ng/ml for serum or saliva cotinine among adults (Pirkle *et al.,* 1996; Caraballo *et al*., 2001, 2004; Society for Research on Nicotine and Tobacco, 2002) and 5.0-11.4 ng/ml saliva or serum for adolescents (McNeill *et al*., 1987; Caraballo *et al*., 2004; Post *et al*., 2005). Optimally, a cut-off is selected in a manner that results in the highest accuracy, defined as the best combination of sensitivity and specificity (Caraballo *et al*., 2001, 2004). However, actual users may have cotinine levels below the cut-off if their most recent use was not recent enough or of sufficient intensity (in terms of units/day) to generate adequate levels of cotinine to exceed the cut-off, and thus be incorrectly classified as deceivers (Dolcini *et al*., 1996; Caraballo *et al*., 2004). Alternatively, some actual non-users of a product (e.g. cigarettes) may be exposed to extremely high doses of secondhand smoke, or they may use other tobacco products or nicotine replacement therapy, and thus may test positive for cotinine. Exposure to secondhand smoke, and use of other tobacco products that are available in a given nation, should be determined by questionnaire assessment and accounted for in validity assessments. In addition, cotinine levels may be influenced by racial/ethnic differences in the rate of nicotine metabolism and intake of nicotine per cigarette smoked (Caraballo *et al*., 1998; Perez-Stable *et al*., 1998; Benowitz *et al*., 2002), suggesting that different cut-offs may be needed for different racial/ethnic groups. Furthermore, the cut-off for pregnant women is lower (e.g. 10 ng/ml) than for the general adult population (Rebagliato *et al.,* 1998; Owen & McNeil, 2001; Society for Research on Nicotine and Tobacco, 2002).

Self-reports from studies with a high demand for abstinence can be biased (Velicer *et al*., 1992; Patrick *et al*., 1994; Benowitz *et al*., 2002). Misclassification of use and non-use has been observed in clinical studies of adult smokers who have been advised to quit and subsequently interviewed about their smoking, often times by persons associated with the intervention. This is particularly true among subjects who have diseases or conditions that would benefit from quitting. For example, it was reported that 15 (65%) of 23

self-reported quitters in a cessation trial of chronic obstructive pulmonary disease patients in the Netherlands misreported use as non-use (Monninkhof *et al*., 2004). In a US study to increase smoking cessation among pregnant women, 49% of self-reported quitters receiving the intervention misclassified use as non-use (Kendrick *et al.,* 1995). In the UK, 11 (22%) of 51 myocardial infarction survivors who had been advised to quit smoking misclassified use as non-use when followed-up during the year after infarction (Sillet *et al*., 1978). In the same report, 40% of subjects in a trial of nicotine gum misclassified their use as non-use.

Population-based surveys, however, are, in general, comprised of people who experience smoking-attributable morbidity at approximately the rate of the general population, are not linked to advice to quit, and administered by interviewers or data collectors who are not known to the respondent. In general, self-reports of current use from surveys are reasonably accurate, providing estimates of prevalence that are comparable to those obtained from use of a biomarker (Pierce *et al*., 1987; Velicer *et al.,* 1992; Patrick *et al.,* 1994; Caraballo *et al*., 2001, 2004; Vartiainen *et al.,* 2002). Data from the surveys used to evaluate the North Karelia project indicate very little misclassification of use as non-use, with no difference in misclassification in North Karelia, where the community-based intervention took place, compared to three other Finnish communities (Vartiainen *et al*., 2002).

However, in cultures in which smoking among women is socially unacceptable, misclassification appears to be more common. Household interviews were conducted on 1403 Southeast Asian adult immigrants who resided in the USA (Wewers *et al.,* 1995). The cotinine-adjusted estimates of current smoking prevalence were substantially higher than those based on self-report for Cambodian females (21.5% versus 6.6%) and Laotian females (10.8% versus 4.2%). In 1992, health surveys were conducted among 1000 adults residing in Pitkäranta in the District of Karelia, Russia and among 2000 adults residing in North Karelia, Finland (Laatikainen *et al*., 1999). The cotinine-adjusted estimates of current smoking prevalence were substantially higher than estimates based only on self-report among women from Pitkäranta (21% versus 10%) than among women from North Karelia (16% versus 13%). The researchers attributed the difference to misclassification of actual use as non-use, most likely because of the social unacceptability of smoking among women in that region of Russia. More recently, concerns were raised about misclassification of use as non-use in population-based surveys conducted in the UK and Poland (West *et al*., 2007). For the UK, cotinine-adjusted prevalence estimates were 2.8 percentage points higher than estimates based on self-report (27.5% versus 24.7%); for Poland, the difference was 4.2 percentage points (41.8% versus 37.6%).

Misclassification of use as non-use is also more likely in household interviews with adolescents, where privacy may be compromised and disclosure is lessened among those who do not want their parents to learn about their behaviour (Turner *et al.,* 1992; US Department of Health and Human Services, 1994; Brittingham *et al*., 1998; Fowler & Stringfellow, 2001; Kann *et al.,* 2002). The prevalence of seven tobacco use behaviours was studied (e.g. lifetime cigarette use, current cigarette use, current smokeless tobacco use, current cigar use) in an experiment that varied mode of administration (paper-and-pencil instrument (PAPI) with computer-assisted self-interview (CASI) and survey setting (school versus home)) (Brener *et al*., 2006). Prevalence differed only for smoking a whole cigarette before age 13 (lower in the PAPI condition) and current smokeless tobacco use (higher in the school setting). Thus, for most of the tobacco-use behaviours measured, home settings can provide prevalence estimates as high as school settings if privacy is increased (both PAPI and CASI afford more privacy than either face-to-face or telephone interviews). It was also demonstrated that when adequate privacy is provided, estimates of cigarette smoking from adolescent surveys conducted in households are similar to those obtained from

surveys conducted in school settings (Gfroerer *et al.*, 1997). Privacy in these studies is afforded by computer-assisted technology, which may not be available in all countries. The four major surveys of adolescents discussed in this Handbook (see Section 4.3) are conducted in schools, which afford even more privacy than homes and provide more efficient venues for data collection.

Self-reports of the number of cigarettes smoked each day appear to be underreported in surveys (Hatziandreu *et al.*, 1989; Section 4.2). Even though cotinine levels increase with increasing number of cigarettes smoked each day (Caraballo *et al.*, 2001; Blackford *et al.*, 2006), survey respondents demonstrate evidence of digit bias towards round numbers (e.g. 10, 15, 20, 30 cigarettes per day) (Klesges *et al.,* 1995), and appear to round down more often than they round up. Comparisons between consumption data and survey-based estimates of consumption should be conducted routinely in countries to provide a crude indicator of the discrepancies between the two sources of information.

Some adolescent survey respondents may indicate they smoke or use smokeless tobacco when they actually do not, perhaps to impress their friends (Cohen *et al.*, 1988; Fowler & Stringfellow, 2001; Stein *et al.,* 2002). However, misclassifying non-use as use appears to be far less common than misclassifying use as non-use (Stein *et al.*, 2002). Adolescent reports that they have smoked during a recent period of time, even when cotinine levels are below threshold values, may still be accurate, because nicotine dosing from infrequent smoking may not result in levels of cotinine that are high enough to exceed the cut-off value (Caraballo *et al.,* 2004, Dolcini *et al.,* 1996). The Centers for Disease Control and Prevention conducted a test-retest study of reporting and found that answers were reasonably stable over a two-week period, with estimates of prevalence being virtually identical (Fowler & Stringfellow, 2001; Brener *et al.*, 1995). The reliability of answers does not prove that they were not distorted on both occasions, but remembering an exaggerated answer is likely more difficult than remembering a true one (Fowler & Stringfellow, 2001).

Methods to enhance validity:

Methodological techniques have been developed to enhance privacy in survey settings, such as having the respondent complete a paper-and-pencil survey form instead of answering a face-to-face interview, which can be overheard (Brittingham *et al.*, 1998); listen to survey questions using headphones connected to a laptop computer, providing answers via the keyboard (Horm *et al.*, 1996; Brener *et al.*, 2006); and respond to questions posed in a telephone interview by pressing the appropriate number button on the key pad instead of replying verbally (Biener *et al.*, 2004). An experiment was conducted to determine if estimates of adolescent drug use obtained from data collected confidentially would differ from those based on data that were collected anonymously (O'Malley *et al.,* 2000). They observed no differences in prevalence estimates, but cautioned that any work conducted without anonymity must convince respondents that all their answers will be kept completely confidential. If a survey respondent believes that the veracity of their self-report will be checked biochemically, then they may be more likely to disclose use (Murray & Perry, 1987; Cohen *et al.*, 1988; Aguinis *et al.*, 1993).

Question wording can also influence the validity of self-report (Babor *et al.*, 1990; Brener *et al.,* 2003; Section 2.2). Survey respondents must first understand a question, interpret it properly, and then encode it into memory. The outputs from this process are then used to search memory and retrieve relevant information, which is evaluated in the decision-making stage of the process. If the information retrieved is considered to be an adequate response, then a response will be generated. If not, then additional retrieval attempts will be made, sometimes involving estimation strategies or adoption of simple rules of thumb that people use to make judgements quickly and efficiently.

If questions are difficult to understand, for example by asking about more than one concept, then the accuracy of response will be compromised. If questions are biased, for example by presenting tobacco use in a negative context,

then answers will also likely be biased. Survey questions must be clear and objective, and constructed in a manner that involves the use of cognitive interviewing techniques, such as those described in Section 2.2.

In an experiment involving the use of three different sets of questions assessing smoking behaviours that held all other conditions constant, researchers obtained similar estimates of adolescent smoking prevalence from the three conditions (Brener *et al*., 2004). Using a convenience sample of 4140 high school students (most were 14-18 years old), approximately equal numbers were randomly assigned to receive questions assessing 14 tobacco use behaviours, based on the actual questions or adapting the question styles of one of these three US surveys: Monitoring the Future Survey, Youth Risk Behaviour Survey, or National Household Survey on Drug Abuse. Questionnaire type was significantly associated with three tobacco-use behaviours: lifetime cigarette use, smoking a whole cigarette before age 13, and purchasing cigarettes at a store or gas station. Nine other measures, including those assessing prevalence of cigarette smoking and smokeless tobacco use, did not vary by questionnaire type. No one questionnaire type proved superior in this experiment. Each set of questions was written in a clear and objective manner.

Question wording can also influence the prevalence estimate obtained depending on what is being measured. Adult respondents to the 1992 National Health Interview Survey (NHIS) who had ever smoked 100 lifetime cigarettes were randomly assigned to be asked, "Do you smoke now?" (the question used prior to 1992) or "Do you now smoke cigarettes every day, some days, or not at all?" (the question used since 1992). Prevalence was 25.6% for those who were asked the first question and 26.5% for those asked the second (Centers for Disease Control and Prevention, 1994a). Including an option on non-daily smoking expanded the range of possible affirmative options, and by doing so provided data on an important behaviour, that of occasional smoking.

The effect of question wording on self-disclosure of smoking in a multiethnic prenatal population in the USA was studied (Mullen *et al*., 1991). Questions about smoking were embedded in a survey instrument assessing multiple risk behaviours. In one condition, subjects were asked "Do you smoke?" and were forced to answer either "yes" or "no." All other subjects were asked, "Which of the following statements best describes your cigarette smoking. Would you say: 1) I smoke regularly now, at about the same amount as before finding out I was pregnant; 2) I smoke regularly now, but I've cut down since I found out I was pregnant; 3) I smoke every once in a while; 4) I have quit smoking since finding out I was pregnant; or 5) I wasn't smoking around the time I found out I was pregnant, and I don't currently smoke cigarettes." The prevalence of smoking was higher in the group given multiple response options (14.0%), compared to the group given the usual question with the dichotomous response categories (9.2%). Most of the women given the multiple choice question reported that they had cut down since learning that they were pregnant, a response option that allows them to disclose their smoking and still display a partially positive image. The researchers estimated that this increase in disclosure would identify an additional 55000 pregnant smokers in the USA each year. In a survey conducted among pregnant women in the UK, cigarette smokers were identified as those who answered "yes" to the question, "Do you smoke at all nowadays?" Approximately 4% of pregnant women misclassified use as non-use (Owen & McNeill, 2001). Widespread adoption of the question used by Mullen and col-leagues might reduce such misclassification.

The overall content of a questionnaire may also influence disclosure. Respondents answering a questionnaire that allows them to portray some positive attributes may be more likely to disclose negative attributes, than if they were answering a questionnaire that only assessed negative attributes (Fowler & Stringfellow, 2001).

In 2002, the Society for Research on Nicotine and Tobacco Subcommittee on Biochemical Verification concluded that the added precision gained by biochemical verification is not

required and may not be feasible in large-scale population-based studies with limited face-to-face contact (Society for Research on Nicotine and Tobacco, 2002). Nevertheless, strategic assessment of validity in situations in which social desirability may lead to substantial underreporting, could be beneficial (Wewers *et al.*, 1995; Laatikainen *et al.*, 1999). In addition, data collected in countries that routinely gather biospecimens for cotinine validation and assessment of exposure to secondhand smoke, could provide a sense of the scope and nature of underreporting, especially as tobacco control progresses and tobacco use becomes increasingly undesirable in a given society.

Issues to consider when comparing different survey estimates

Surveillance and evaluation systems will provide comparable estimates of tobacco use behaviours to the extent that they use similar methods. The factors that influence validity (e.g. assurance of privacy and that answers will remain completely confidential, question wording, social desirability) will influence estimates of prevalence and thus comparisons between surveys. Factors that can influence prevalence estimates in ways that do not influence validity are described below.

Definition of a user:

Differing definitions of a "user" will often yield differing estimates of prevalence of use. For example, in a country where multiple forms of tobacco are available, as in India and the USA, a survey providing an estimate of a *tobacco use* would result in a higher estimate of prevalence than one that only reports on the prevalence of *tobacco smoking*. Similarly, an estimate of *cigarette smoking* prevalence would be lower than estimates of tobacco use and of tobacco smoking. In the same way, estimates of *current daily smoking* would be lower than estimates of *current smoking*, which include both daily and non-daily smoking.

Sample frame:

The sample frame of a survey can influence the prevalence estimates generated. For example, prevalence could differ substantially for surveys of persons aged 15 years and older, aged 25 years and older, and 25 to 64 years old. Likewise, a frame drawn only from major metropolitan areas in a given country would likely produce substantially different prevalence estimates than if the entire population were sampled. Each of the estimates from the sample frames discussed here could be valid for the population covered by the respective sample frame. Thus, knowledge of each survey's sample frame is important when making comparisons across surveys.

Another sample frame issue deals with telephone coverage. Telephone surveys are frequently conducted in developed countries. The major advantage of such surveys is that they are less expensive to conduct than household interviews. Telephone surveys are generally not conducted in developing countries, where coverage does not permit the drawing of a representative sample. In developed countries, however, the increasing prevalence of adults who own a wireless telephone, but live in a household with no landline telephone, presents a potential for bias, because sample frames for telephone surveys are drawn from numbers for landline telephones. According to data from the 2004 and 2005 US National Health Interview Survey (NHIS), approximately 1.7% of adults lived in households that did not have any telephone service, 5.6% of adults lived in households with only wireless telephones, and 92.8% of adults lived in households with landline telephones (Blumberg *et al.,* 2006). The prevalence of cigarette smoking was 19.7% (95% CI: 19.2-20.2) among adults living in households with landline telephones, 32.9% (95% CI: 30.9-35.0) among adults in households with only wireless telephones, and 36.9% (95% CI: 33.4-40.3) among adults in households with no telephone service. Thus, all other things being equal, the prevalence of cigarette smoking that would have been estimated from a telephone survey, that only reached households with landline telephones, would have been 19.7%, whereas the prevalence in all households in the NHIS was 20.9%, a difference of 1.2 percentage points ($P < 0.05$). Telephone surveys provide valu-

able information. Rates of coverage will likely vary across nations. The small difference in cigarette smoking prevalence estimates seen in the USA suggest that comparisons of prevalence estimates from telephone and household surveys should consider the possible influence of coverage bias.

Samples for surveys of adolescents are drawn either from school-based frames, providing access to enrolled students, or from household lists and subsequent enumerations of household members. Only household frames provide access to school dropouts, who are more likely to smoke cigarettes than students of the same age (Gfroerer *et al.,* 1997). This issue poses greater concern for older (i.e. ages 16-17 years) adolescents than for their younger counterparts, who are less likely to have dropped out of school. Another comparability issue is that household surveys may not report data for an age group that is comparable to one found in a school survey. For example, if a household survey reports estimates for young people who are 12-17 years old, and a school survey reports estimates for students enrolled in grades 9-12 (most of whom are 14-18 years old), then the school survey will likely have higher prevalence estimates simply because there are no 12-13 year olds enrolled in schools in this frame, and the household age group does not include 18 year olds. Consumers of survey data should consider these and other factors when comparing data from school and household surveys.

Editing procedures:

Surveys that are administered via self-administered questionnaires, such as the youth surveys described in Section 4.3, require decision rules for dealing with inconsistent answers. The effects of five approaches for handling such inconsistencies in the 1998 Florida Youth Tobacco Survey were described (Bauer & Johnson, 2000). The approaches ranged from doing nothing, which ignored inconsistencies and analyzed each item as a separate entity, to a "preponderance" approach, which evaluated each record and assigned values based on the weight of the evidence for each respondent. The cigarette smoking prevalence estimates generated from these approaches ranged from 25.6% (95% CI: 24.1-27.1) to 29.7% (95% CI: 28.2-31.2). Boys exhibited more inconsistencies and therefore more variability across approaches. While recognizing the impossibility of discerning which approach is the most valid, the authors suggested that editing procedures be described when findings are reported. Approaches for handling inconsistencies can influence prevalence estimates and survey comparability (Brittingham *et al.,* 1998; Bauer & Johnson, 2000).

Type of survey:

Recent reports indicate that prevalence estimates obtained from surveys in California (Cowling *et al*., 2003) and New Hampshire (Ramsey *et al*., 2004) in the USA are lower in surveys with a tobacco focus than in general health surveys. The phenomenon was studied using a factorial design and concluded, after a series of multivariate analyses, that the introduction to the tobacco survey cued some people, mainly women, who didn't want to spend the time on the survey, to misclassify themselves as non-users (Cowling *et al*., 2003). The researchers argued that the social stigmatization of tobacco use in California may have contributed to the misclassification bias they observed.

Type of parental consent in school-based surveys of adolescents:

In most countries, letters are sent home notifying parents that their children will participate in a survey (parental notification). In some countries, such as the USA and Australia, two types of parental permission are required for school-based survey research. In both systems, a letter is sent to parents describing the upcoming survey research project and requesting their child's participation. In active parental permission, a form must be returned, signed by a parent, granting the child permission to participate. If no signed form is returned, disapproval is assumed. In passive permission, parents send back a signed form only if they do not want their child to participate. If no form is returned, parental approval is assumed. In the USA, selected state and municipal governments require active

permission. Three US reports have noted that estimates of tobacco use are lower when active parental permission is required (Severson & Ary, 1983; Dent *et al.*, 1993; Anderman *et al.*, 1995). It is suggested that active permission laws exclude high risk students because they are less likely to return signed permission forms. Differences were not observed in ever smoking or smoking during the previous week in a study of active versus passive consent conditions in Australia (White *et al.,* 2004).

An analysis of the 2001 Youth Risk Behaviour Survey (YRBS) data was undertaken to determine if type of parental consent was related to the magnitude of estimates for 26 behaviours, including lifetime cigarette smoking, current cigarette smoking, and current smokeless tobacco use (Eaton *et al.*, 2004). Of 13195 eligible students, 65% lived in passive conditions. In passive condition schools, 86.7% of sampled students participated; 77.3% of students in active condition schools did so. The difference was due to the 9.5% of students in the active condition who did not return a permission form. Type of consent did not influence any of the tobacco measures; in fact, it was related to only two of the 26 behaviours measured. The conclusion was that the requirement for active consent will not influence prevalence estimates if participation rates are sufficiently high (Eaton *et al.*, 2004). It was also argued that the anonymity offered by the YRBS might have lessened any concerns students had about their parents' negative attitudes about certain risk behaviours and facilitated disclosure. Thus, comparisons of estimates from school surveys in various countries should assess the degree to which active consent is required and the participation rate in each condition.

Response rates:

Concern has been raised about the effects of declining response rates in telephone surveys, especially in the USA. As the US rates declined in the 1990s, no differences in the degree of representation in samples of population subgroups were observed (Biener *et al.*, 2004). The researchers also compared cigarette smoking prevalence estimates from telephone surveys conducted in Massachusetts and California, where response rates dropped substantially, with those from the Tobacco Use Supplement to the Current Population Survey (TUS-CPS), in which response rates dropped only very slightly and were substantially higher in 1998-1999 (76%-81% in the TUS-CPS versus 69% in Massachusetts and 51% in California). The smoking prevalence estimates obtained from the Massachusetts and California surveys remained reasonably close (as judged by overlapping confidence intervals) to those from the TUS-CPS, with no evidence of an increasing disparity over time.

Despite the findings from this study, researchers should work diligently to maximize response rates, and continue to monitor response rates, sample characteristics, and prevalence estimates across surveys with differing response rates to identify variables that might compromise comparisons.

Survey-based measures of tobacco use behaviours

A general outline of the variables used to monitor the natural history of tobacco use is presented in Table 3.1. A description of detailed question items for almost every component of the process, and some commentary on each, are provided in Tables 3.2 through 3.18. Intention to try (I.a. in Table 3.1) and intention to quit (IV.a. in Table 3.1) are discussed in Section 3.2. The methods used in cessation attempts (IV.b.iii. in Table 3.1) are discussed in Section 5.7. Topography (as an indicator of smoke intake) (III.e. in Table 3.1) is discussed in the text below; however, no survey items are recommended for this topic, as questionnaire assessments of smoking topography have not been shown to be valid.

Tables 3.2 through 3.18 list questions relevant for each topic that is either used in the cross-national surveys described in Section 4.3, or in country-specific surveys. The latter are added in instances where they supplement the items used in the cross-national surveys. In reliability assessments shown in the tables, kappa statistics of 61-80% were considered substantial and 81-100% were almost perfect (Brener

Construct	Construct I.b. on Table 3.1 (Initial Trial)
Measure	"On how many occasions (if any) during your lifetime have you smoked cigarettes?" Number of occasions: 0, 1-2, 3-5, 6-9, 10-19, 20-39, 40 or more (ESPAD) "How old were you when you first tried a cigarette?" I have never smoked cigarettes; 7 years old or younger; 8 or 9 years old; 10 or 11 years old; 12 or 13 years old; 14 or 15 years old; 16 years old or older (GSHS) "Have you ever tried or experimented with cigarette smoking, even one or two puffs?" (GYTS) "Have you ever smoked tobacco?" (at least one cigarette, cigar or pipe) (HBSC)
Sources	ESPAD, GSHS, GYTS, HBSC
Validity	Face validity. Kappa for ever use of cigarettes was 83.8% in CDC 14-day reliability study among high school students (Brener *et al.*, 1995). 81.5% agreement in a two year study (Shillington & Clapp, 2000). 92.3% of baseline ever users reported consistently at follow-up survey, with consistency decreasing with increasing time between assessments (Huerta *et al.,* 2005).
Variation	Items are adaptable for assessments of other tobacco products. For example, a survey could ask, "On how many occasions (if any) during your lifetime have you used smokeless tobacco?" Number of occasions: 0, 1, 2-3, 4-9, 10-19, 20-39, 40 or more
Comments	This variable is assessed mostly in youth surveys. The only cross-national adult survey which conceptually can indicate ever use is the GATS, which asks non-current users: "In the past, have you smoked tobacco (cigarettes, cigars or pipes) on a daily basis, less than daily, or not at all?"
Definitions	Ever users have tried one or more smoke or smokeless tobacco products. Never users have not tried tobacco, even the least amount asked about. Definitions more specific to product type(s) can be employed (e.g. ever smoker, ever cigarette smoker, ever user of smokeless tobacco, ever user of betel quid).

GYTS: Global Youth Tobacco Survey
HBSC: Health Behaviour of School-aged Children
ESPAD: European School Survey Project on Alcohol and Other Drugs
GSHS: Global School Health Survey
GATS: Global Adult Tobacco Survey
CDC: Centers for Disease Control and Prevention

Table 3.2 Initial Trial - Ever Use of Cigarettes or Smoked Tobacco

et al., 1995). Also, intraclass correlation coefficients (ICC) of 0.75 and higher were considered excellent, and 0.60 to 0.74 were considered good (Johnson & Mott, 2001). Most of the measures are listed in terms of smoking behaviour. Modifications of each item can be made for smokeless tobacco use.

Initial trial:

This construct distinguishes persons who have never used from those who have ever used tobacco (Table 3.2). The proportion of young people who have never tried a cigarette is one of the Center for Disease Control and Prevention's (CDC) key outcome indicators (Starr *et al.*, 2005). Reducing the number of people who ever try tobacco will reduce the number who become established users (US Department of Health and Human Services, 1994; Starr *et al.*, 2005). Best measured in school surveys of adolescents, initial trial can be assessed for whichever tobacco

products are of most relevance in a particular country. Trends in this measure have been studied for more than 30 years in the USA, where lifetime use of cigarettes among high school seniors (i.e. 12th grade students, the vast majority being 17-18 years old) was 73.6% in 1975 and 50% in 2005 (Johnston *et al.*, 2006). Cross-national findings on initial use have been reported in several reports (Warren *et al.*, 2000; Global Youth Tobacco Survey Collaborative Group, 2002; Godeau *et al.*, 2004; Hibell *et al.*, 2004; Global Tobacco Surveillance System Collaborating Group, 2005; White & Hayman, 2006). Here we define a "trier" as someone who has tried smoking, but has only taken one or more puffs, but never a whole cigarette/cigar/pipe, or as someone who has tried smokeless tobacco, but only on one occasion (Table 3.3).

The age of first use is another CDC key outcome indicator (Starr *et al.*, 2005). The younger people are when they start using tobacco, the more likely they are to use it as adults (US Department of Health and Human Services, 1994). Trends over time in average age or grade of first use have been reported (Kopstein, 2001; Johnston *et al.,* 2006). Measures of actual age of first use have been used to calculate the incidence of initiation of first use (Centers for Disease Control and Prevention, 1998; Kopstein, 2001). The average age of first use varies across countries, likely reflecting the influence of media and of cultural values (Warren *et al.,* 2000; Global Youth Tobacco Survey Collaborative Group, 2002; Global

Construct	Construct I.b. and I.c. on Table 3.1 (Initial Trial and Experimentation)
Measure	"How many cigarettes have you smoked in your entire life?" None; 1 or more puffs, but never a whole cigarette; 1 cigarette; 2 to 5 cigarettes; 6 to 15 cigarettes (about ½ pack total); 16 to 25 cigarettes (about 1 pack total); 26 to 99 cigarettes (more than 1 pack but less than 5 packs); 100 or more cigarettes (5 or more packs) (GYTS – OPTIONAL)
Source	GYTS
Validity	Face validity. 10-18 year old US smokers who had smoked 20-98 lifetime cigarettes were more likely to report that they smoked because it "relaxes or calms" them and because "it's really hard to quit" than were smokers who had smoked fewer than 20 lifetime cigarettes (Centers for Disease Control and Prevention, 1994a).
Variation	Items are adaptable for assessments of other tobacco products. For example, a survey could ask, "On how many occasions (if any) during your lifetime have you used smokeless tobacco?" Number of occasions: 0, 1, 2-3, 4-9, 10-19, 20-39, 40 or more The parenthetical examples of the number of packs listed in the item above for cigarettes apply only in countries in which there are 20 cigarettes in each package.
Comments	Definitions for cigarette smoking are based on Choi *et al.*, 2001.
Definitions	A trier is someone who has tried smoking, but has only taken a few puffs or someone who has tried smokeless tobacco, but only once. An experimenter is someone who has smoked more than a few puffs, but fewer than 100 cigarettes. For other tobacco products, the US National Center for Health Statistics uses cut-offs of from 1-49 cigars or pipes full of tobacco or having used smokeless tobacco on from 1-19 occasions.

GYTS: Global Youth Tobacco Survey

Table 3.3 Trial versus Experimentation

Tobacco Surveillance System Collaborating Group, 2005). Table 3.4 describes the construct "Age of First Use."

Discontinuation after initial trial:

Some young people will try tobacco, for example, by taking a few puffs on a cigarette, and then never use again. Tobacco control policies aim first to prevent initial trial and, if initial use has occurred, to prevent progression beyond such use. Researchers used one month with or without use to distinguish "recent" from "non-recent" experimenters (Choi *et al.*, 2001). However, approximately three in 10 non-recent experimenters, according to their definition, progressed to established use. The question recommended in Table 3.5 permits use of other time periods after initial trial. Three months since initial use can be used to define former triers. This strategy, while somewhat arbitrary, is based on the assumption that triers who have not used for at least three months, would be less likely to progress to established user than

Construct	Construct I.b. on Table 3.1 (Inital Trial)
Measure	"When (if ever) did you first do each of the following things?" A) Smoke your first cigarette? Never; 9 years old or less; 10 years old; 11 years old; 12 years old; 13 years old; 14 years old; 15 years old; 16 years or older (ESPAD) "How old were you when you first tried a cigarette?" I have never smoked cigarettes; 7 years old or younger; 8 or 9 years old; 10 or 11 years old; 12 or 13 years old; 14 or 15 years old; 16 years old or older (GSHS) "How old were you when you first tried a cigarette?" I have never smoked cigarettes; 7 years old or younger; 8 or 9 years old; 10 or 11 years old; 12 or 13 years old; 14 or 15 years old; 16 years old or older (GYTS) "At what age did you first do the following things? Smoke a cigarette:" Never, ___ (write in age). (HBSC)
Sources	ESPAD, GYTS, GSHS, HBSC
Validity	Face validity. Kappa for smoking first whole cigarette before age 13 years was 68.1% in CDC 14-day reliability study among high school students (Brener *et al.,* 1995). Intraclass correlation coefficient (ICC) was good (range = .637 - .666) in three tests of children and moderate (0.517) in a fourth in a two year reliability study (Johnson & Mott, 2001). The ICC was 0.73 for males and 0.76 for females in an Israeli study (Huerta *et al.*, 2005). Forward telescoping (producing older estimates of age of first use upon re-interview) has been observed (Shillington & Clapp, 2000; Johnson & Mott, 2001).
Variation	Items are adaptable for assessments of other tobacco products.
Comments	The NSDUH asks adolescents and adults, "How old were you the first time you smoked part or all of a cigarette?" (http://oas.samhsa.gov/nsduh.htm). This measure has been used to assess incidence of initiation (Centers for Disease Control and Prevention, 1998); NSDUH even assesses month of first use in recent initiators (http://www.oas.samhsa.gov/2k4/season/season.htm).

ESPAD: European School Survey Project on Alcohol and Other Drugs
GSHS: Global School Health Survey
GYTS: Global Youth Tobacco Survey
HBSC: Health Behaviour of School-aged Children
CDC: Centers for Disease Control and Prevention
NSDUH: US National Survey on Drug Use and Health

Table 3.4 Age of First Use

would those abstinent for less than three months.

Experimentation:

Experimentation occurs when someone progresses beyond initial trial. Experimentation with cigarettes can be distinguished from initial trial and from established use with the question recommended in Tables 3.3 and 3.6. Experimenters are those who have consumed from 1-99 cigarettes. Regarding the use of other tobacco products, experimentation can be operationalised as smoking from 1-49 cigars or pipes full of tobacco, or having used smokeless tobacco on from 2-19 occasions. These are somewhat arbitrary cut-offs; the US National Center for Health Statistics uses 50 cigars, 50 pipes full of tobacco, and use of smokeless tobacco on at least 20 occasions to measure established use in a manner similar to the 100 cigarette question. Indicators of nicotine dependence have been observed during the experimentation process (Centers for Disease Control and Prevention, 1994b; DiFranza *et al.,* 2002b; O'Loughlin *et al*., 2003).

Discontinuation of experimentation:

Another goal of tobacco control is to prevent the progression from experimentation to established use. As discussed above, a cut-off of three months of abstinence since experimenting can be used to define former experimenters (see Table 3.5).

Transition to established use:

Young people who have become established users are, compared to those who have not, at far greater risk of continuing to smoke as adults (US Department of Health and Human Services, 1994; Choi *et al*., 2001). Preventing progression to established use is a goal of tobacco control. CDC has identified the proportion of young people who have smoked 100 cigarettes or more during their lifetimes as a key outcome indicator for evaluating comprehensive tobacco control programmes (Starr *et al*., 2005). Similar indicators for other tobacco products are recommended in Table 3.6. Several other measures of transition have been described as well (Johnston, 2001).

Construct	Construct I.b.i and I.c.i. on Table 3.1 (Discontinuation)
Measure	"When was the last time you smoked a cigarette, even one or two puffs?" I have never smoked a cigarette; today; not today, but some time during the past week; not in the past week, but some time in the past month; 2-3 months ago; 4-6 months ago; 7-12 months ago; 1 or more years ago (GYTS – OPTIONAL)
Source	GYTS
Validity	Face validity. In one study, non-recent experimenters (those experimenters who had not smoked within the previous 30 days) were less likely to progress to established smoking than were current experimenters (Choi *et al*., 2001).
Variation	Items are adaptable for assessments of other tobacco products.
Definitions	A *former trier* is someone who has smoked only a few puffs or who has tried smokeless tobacco only once who has not used it for ≥ 3 months. *A former experimenter* is someone who has experimented (defined in Table 3.3) and has not smoked/used tobacco for ≥ 3 months.

GYTS: Global Youth Tobacco Survey

Table 3.5 Time Since Last Use Among Triers or Experimenters

Construct	Construct II. on Table 3.1(Transition to established use)
Measure	"How many cigarettes have you smoked in your entire life?" None; 1 or more puffs, but never a whole cigarette; 1 cigarette; 2 to 5 cigarettes; 6 to 15 cigarettes (about ½ pack total); 16 to 25 cigarettes (about 1 pack total); 26 to 99 cigarettes (more than 1 pack but less than 5 packs); 100 or more cigarettes (5 or more packs) (GYTS – OPTIONAL) "Have you smoked 100 cigarettes or more in your lifetime?" (ITC) "Have you smoked at least 100 cigarettes in your entire life?" (NHIS, BRFSS, NSDUH, ATS, TUS-CPS)
Sources	GYTS, ITC, NHIS, BRFSS, NSDUH, ATS, TUS-CPS
Validity	Evidence of utility – predictive validity. Adolescents who have smoked at least 100 lifetime cigarettes are more likely to be established smokers in the future than those who have not (Choi *et al*., 2001).
Variation	Items are adaptable for assessments of other tobacco products. "On how many occasions (if any) during your lifetime have you used smokeless tobacco?" Number of occasions: 0, 1, 2-3, 4-9, 10-19, 20-39, 40 or more
Comments	Having ever smoked 100 cigarettes is considered "established" use (Choi *et al*., 2001; Starr *et al*., 2005). It is a useful measure because it can be used as a marker for a threshold even for never daily users. However, some people have difficulty understanding the concept of having ever smoked a total of 100 lifetime cigarettes. For other tobacco products, the use of $\geq$ 50 cigars or pipes full of tobacco or having used smokeless tobacco on $\geq$ 20 or more occasions can be used as cut-offs to define established use.

GYTS: Global Youth Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
NHIS: US National Health Interview Survey
BRFSS: US Behavioural Risk Factor Surveillance System
NSDUH: US National Survey on Drug Use and Health
ATS: US Adult Tobacco Survey
TUS-CPS: US Tobacco Use Supplement to the Current Population Survey

Table 3.6 Threshold for Transition to Regular Use

Ever daily versus never-daily:

In the USA in 1991, approximately 7.5% of established smokers had never smoked on a daily basis (Husten *et al*., 1998). Among all established smokers, never daily smoking was more common among non-Whites (range = 12-17%) than among Whites (6%); among current smokers, never daily smoking was also more common among non-Whites (range = 11-17%) than among Whites (4%).

The average age of first daily use can vary among ethnic groups within a country and over time (Centers for Disease Control and Prevention, 1991). Compared with younger age of first daily use, starting at an older age has been associated with slightly lower rates of subsequently developing tobacco-attributable disease (US Department of Health and Human Services, 2004). Description of ever daily use constructs and age of first daily use are found in Tables 3.7 and 3.8.

Current use:

Current use is influenced primarily by rates of initiation and quitting, as well as by mortality, and to a far lesser extent, immigration into and emigration out of a given population. Current use is the most important construct because of its importance as an outcome variable in policy evaluation studies. CDC rates it a key outcome indicator (Starr *et al*., 2005).

Each of the seven surveys described in Section 4.3 measures current use (Table 3.9). In

three (European School Survey Project on Alcohol and Other Drugs (ESPAD), Global School Health Survey (GSHS), Global Youth Tobacco Survey (GYTS)) of the four surveys of young people, a current user is someone who used tobacco at least once during the previous 30 days (month) (Warren *et al*., 2000, 2006; Hibell *et al*., 2004; WHO, 2007a). In the Health Behaviour of School-aged Children (HBSC) survey, a current user is someone who uses either daily or weekly (Godeau *et al.,* 2004; Hublet *et al*., 2006). Current use is defined slightly differently in the adult surveys. In the Global Adult Tobacco Survey (GATS) and the STEPwise Approach to Chronic Disease Factor Surveillance (STEPS) survey, a current smoker is someone who currently smokes tobacco products daily or less than daily. GATS and STEPS can distinguish between current daily and current non-daily smoking (Table 3.9). GATS can also classify current non-daily smokers as ever daily or never daily smokers. The International Tobacco Control Policy Evaluation Survey (ITC) classifies current cigarette smokers as those who had ever smoked ≥ 100 lifetime cigarettes who currently smoke daily, weekly, or monthly.

Trends in and patterns of current use have been reported in numerous reports and publications (US Department of Health and Human Services, 1994,1998, 2001; Warren *et al*., 2000; Kopstein, 2001; Giovino, 2002; White & Hayman, 2006). The WHO Global InfoBase documents prevalence of current use of various indicators, including current smoking, current daily smoking, and current tobacco use for countries throughout the world

Construct	Construct II.a. on Table 3.1 (Ever daily and never daily)
Measure	"When (if ever) did you first do each of the following things? B) Smoke cigarettes on a daily basis:" Never; 9 years old or less; 10 years old; 11 years old; 12 years old; 13 years old; 14 years old; 15 years old; 16 years or older (ESPAD) "Have you ever smoked cigarettes daily, that is, at least one cigarette every day for 30 days?" (NYTS) "In the past, have you smoked tobacco (cigarettes, cigars or pipes) on a daily basis, less than daily, or not at all?" (GATS) "In the past, did you ever smoke daily?" (STEPS)
Sources	ESPAD, NYTS, GATS, STEPS
Validity	Face validity. Kappa for ever daily use was 86.6% in CDC 14-day reliability study among high school students (Brener *et al*., 1995).
Variation	In GATS, current non-daily smokers are asked, "Have you smoked tobacco daily in the past?" Items are adaptable for assessments of other tobacco products.
Comments	The prevalence of never daily smoking among adult smokers in the USA was documented (Husten *et al*., 1998).
Definitions	An ever daily user is someone who has ever smoked tobacco or used smokeless tobacco on a daily basis. A never daily user has never smoked tobacco or used smokeless tobacco on a daily basis.

ESPAD: European School Survey Project on Alcohol and Other Drugs
NYTS: National Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance
CDC: Centers for Disease Control and Prevention

Table 3.7 Ever daily versus Never Daily Use

Construct	Construct II.a. on Table 3.1 (Ever daily and Never Daily)
Measure	"When (if ever) did you first do each of the following things? Smoke cigarettes on a daily basis:" Never; 9 years old or less; 10 years old; 11 years old; 12 years old; 13 years old; 14 years old; 15 years old; 16 years or older (ESPAD) "How old were you when you first started smoking daily?" (GATS, STEPS)
Sources	ESPAD, GATS, STEPS
Validity	Face validity. Kappa for first smoking daily before age 13 years was 71.8% in CDC 14-day reliability study among high school students (Brener *et al.,* 1995). ICC was excellent for adults' assessments of age of first daily use (.815) in a two year reliability study (Johnson & Mott., 2001). Forward telescoping (producing older estimates of age of first daily use upon re-interview) has been observed (Johnson & Mott., 2001).
Variation	Items are adaptable for assessments of other tobacco products.
Comments	The NSDUH asks adolescents and adults, "How old were you when you first started smoking every day?" (http://oas.samhsa.gov/nsduh.htm). This measure has been used to assess incidence of initiation of daily use (Centers for Disease Control and Prevention, 1998). Measures like this have been used to calculate incidence of initiation of cigarette smoking (Pierce *et al.*, 1994; Pierce & Gilpin, 1995; Centers for Disease Control and Prevention, 1998).

ESPAD: European School Survey Project on Alcohol and Other Drugs
GATS: Global Adult Tobacco Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance
CDC: Centers for Disease Control and Prevention
NSDUH: US National Survey on Drug Use and Health

Table 3.8 Age at first daily use

(http://www.who.int/ncd_surveillance/infobase/web/InfoBaseCommon).

Frequency of use:

Frequency of use refers to the number of days when tobacco is used during a given time period (e.g. the previous seven days or the previous 30 days). Frequency of use is often dichotomized as either current daily or current non-daily use (Table 3.9). In the USA, current non-daily smoking is more common among African Americans and Hispanics than it is among non-Hispanic Whites (US Department of Health and Human Services, 1998). Overall, current non-daily smoking remained stable at about 18-19% of all current smokers from 1993 to 2004 (Trosclair *et al.*, 2005).

In surveys of young people, current frequent users are those who smoked on $\geq$ 20 or more of the previous 30 days. Frequency of use is a predictor of quitting (with more frequent use associated with a lower probability of subsequent quitting than less frequent use) (Hyland *et al.,* 2004).

Type of product used:

It is important to measure the type of product consumed, particularly in countries, such as India, where there exists a variety of commonly used forms of tobacco products. The variety of forms available, and the possibility of switching or multiple concurrent uses may influence the probabilities of quitting and of disease risk. Country-specific lists of products to be monitored should be incorporated into each country's survey. Examples of items used in the various cross-national surveys are provided in Table 3.10.

Per capita consumption (by weight) of various tobacco products is often documented by government agricultural agencies (Capehart, 2007). A useful rule of

Construct	Constructs III. and III.a. on Table 3.1 (Current use)
Measure	**Surveys of Youth** "How frequently have you smoked cigarettes during the LAST 30 DAYS?" Not at all; less than 1 cigarette per week; less than 1 cigarette per day; 1-5 cigarettes per day; 6-10 cigarettes per day; 11-20 cigarettes per day; more than 20 cigarettes per day (ESPAD) "During the past 30 days, on how many days did you smoke cigarettes?" 0 days; 1 or 2 days; 3 to 5 days; 6 to 9 days; 10 to 19 days; 20 to 29 days; all 30 days (GSHS) "During the past 30 days (one month), on how many days did you smoke cigarettes?" 0 days; 1 or 2 days; 3 to 5 days; 6 to 9 days; 10 to 19 days; 20 to 29 days; all 30 days (GYTS) "Do you smoke now?" Not at all; occasionally, but less than once a month; some time each month, but less than one cigarette per week; sometime per week, but less than one cigarette per day; every day at least one cigarette? (GYTS – OPTIONAL) "How often do you smoke at present?" Every day; at least once a week, but not every day; less than once a week; I do not smoke (HBSC) **Surveys of Adults** "Do you currently smoke tobacco (cigarettes, cigars or pipes) on a daily basis, less than daily, or not at all?" (GATS) "Do you smoke every day, less than every day, or not at all?" (including factory-made cigarettes or hand-rolled cigarettes). NON-DAILY SMOKERS ARE ASKED: "Do you smoke at least once a week?" THOSE WHO ANSWER NO ARE ASKED: "Do you smoke at least once a month?" (ITC) "Do you currently smoke any tobacco products, such as cigarettes, cigars, or pipes?" IF YES: "Do you currently smoke tobacco products daily?" (STEPS)
Sources	ESPAD, GSHS, GYTS, HBSC, GATS, ITC, STEPS
Validity	Evidence of utility. Self-reports of current use have been shown to be reasonably valid for adults and youths, when adequate privacy is afforded (Turner *et al.*, 1992; Velicer *et al.,* 1992; Patrick *et al.*, 1994; US Department of Health and Human Services, 1994; Gfroerer et al., 1997; Brittingham *et al.*, 1998; Caraballo *et al.*, 2001; Fowler & Stringfellow, 2001; Kann *et al.*, 2002; Caraballo *et al.*, 2004; Brener *et al.*, 2006). Kappa for smoking on ≥ 14 days during the previous 30 days was 80.1% in CDC 14-day reliability study among high school students (Brener *et al.*, 1995). Evidence indicated that for persons aged ≥ 18 years, current smoking prevalence estimates based on proxy reports are virtually identical to those based on self-report (Gilpin *et al.*, 1994).
Variation	Items are adaptable for assessments of other tobacco products.
Definitions	**Among Youth**: A current user is someone who used tobacco at least once during the previous 30 days (month). A current frequent user is someone who used tobacco on > 20 of the previous 30 days. **Among Adults**: A current user is someone who consumes tobacco daily or less than daily (GATS, STEPS) or someone who consumes tobacco daily or less than daily during the previous month (ITC). A current daily user is someone who reports using on a daily basis. **Among both Youth and Adults**: Frequency refers to the number of days smoked each month.

Table 3.9 Current Use (Daily versus Non-Daily)

Comments	Comparisons of adolescent prevalence estimates with those of adults can be problematic. For example, estimates of current use among adolescents are often considerably higher than those among adults. However, adolescents who smoke generally do so on fewer days each month than do adult smokers. Ideally, comparisons of use among youth and adults would be made with a measure of the number of days smoked during the previous 30 days (e.g. ≥ 20 of 30 days). In countries where adult surveys do not measure the number of days smoked out of the previous 30 days, then comparing adult prevalence of current use with the prevalence of current frequent use among adolescents would be preferred to comparisons of past month use, because the vast majority of adult users consume tobacco on ≥ 20 of the previous 30 days. Some countries measure use during the previous week. Comparisons of weekly use among adolescents and adults would provide more comparable estimates than past month use.

ESPAD: European School Survey Project on Alcohol and Other Drugs
GSHS: Global School Health Survey
GYTS: Global Youth Tobacco Survey
HBSC: Health Behaviour of School-aged Children
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance
CDC: Centers for Disease Control and Prevention

Table 3.9 Current Use (Daily versus Non-Daily)

thumb is that when the amount of tobacco consumed in a particular product (e.g. snuff) comprises less than 1% of total tobacco consumed, then use of that product need not be assessed in surveys. Exceptions to that rule may occur when use of a product that is rarely consumed in the overall population is more common among a subgroup of the population. In the USA, for example, the use of bidis is rare in the adult population, but of concern among young people (National Youth Tobacco Survey (NYTS) data, US National Survey on Drug Use and Health (NSDUH) data).

Brand used:

The prevalence of use of specific brands among users of a particular product type (e.g. manufactured cigarettes) reflects the influence of both marketing campaigns and product design (Centers for Disease Control and Prevention, 1994c; Tomar *et al.,* 1995; Slade, 2001; Cummings *et al*., 2002a; Wayne & Connolly 2002; Carpenter *et al.,* 2005; Lewis & Wackowski, 2006). Tobacco control practitioners can use this information to implement policies (e.g. counter-marketing campaigns, tobacco product regulation) designed to reduce overall use. Survey-based measures of brand used are presented in Table 3.11; measures of brand switching are described in Table 3.12.

Sub-brand characteristics (e.g. strength, flavoring, length) are often determined by either asking for the name of the specific brand purchased or asking the name of a brand family, followed by each of several possible sub-brand characteristics (Table 3.11). Strength has often been described by industry terms such as “light” and “mild.” Because these terms are misleading (National Cancer Institute, 2001), they have been banned in a number of countries (e.g. European Union countries, Australia) and replaced either by other terms or specific color schemes that indicate strength based on machine-measured yields. All of these indicators are still misleading, since the tests used to determine strength do not reflect actual human exposure (National Cancer Institute, 2001; Hammond *et al.,* 2006b). Thus, it is important to capture the extent of use of these terms, either via survey-based questions (Table 3.11), or via documentation of what is on the actual package.

Detailed measurement of information about tobacco product packaging is important in order to determine the variant of product type used, movement between price sectors, and, potentially, to assess the use of tobacco from illicit sources. Interviewers can either collect empty packages or

take digital photographs of a given respondent's current pack. Package characteristics to document include: brand name, strength, flavoring, length, pack type (hard pack versus soft pack), package color, color in words (e.g. Silk Cut Silver, Silk Cut Purple), filter (e.g. non-filter, charcoal [if designated]), UPC code, number of cigarettes per pack, constituents measured and levels, text, warning label(s) (words, picture [if applicable], and location[s]), and the presence or absence of a tax stamp.

In addition to survey based measures, governments should make available to researchers and policy makers sub-brand-specific sales data on a region-specific basis. This will allow researchers to better document the influence of tobacco product marketing practices.

Intensity of use:

Intensity of use reflects the average number of cigarettes, cigars, or pipes full of tobacco smoked each day for daily smokers, or on the days during which the respondent smoked for non-daily smokers. Selected questionnaire items used to assess intensity are listed in Table 3.13. Intensity decreases following the implementation of smoke-free policies (Fichtenberg & Glantz, 2002a; Section 5.2) and price increases (Chaloupka *et al.,* 2001; Warner, 2006; Section 5.1). Intensity is inversely associated with the probability that a respondent will quit (Hyland *et al.*, 2004), and is directly related to the probability of developing a tobacco-attributable disease (US Department of Health and Human Services, 2004; IARC, 2004).

Smoke intake:

The intake of smoke from a cigarette is generally determined in laboratory studies of smoking topography, which assess how cigarettes are smoked. Variables measured include the number of puffs taken per cigarette, the duration of each puff, inter-puff interval, puff volume, the draw rate of each puff, the unsmoked butt length, and the amount of obstruction of filter ventilation holes (Pechacek *et al.*, 1984). Unfortunately, questionnaire assessments of this construct have not proven to be valid. Two alternative techniques have been developed that estimate smoke intake from the study of cigarette filter butts: one measures the amount of solanesol, a naturally occurring component of tobacco that is deposited during smoking in the cigarette filter butt (Watson *et al.*, 2004a); and the other studies the staining pattern on filter butts as a proxy measure for total smoke volume (O'Connor *et al.*, 2005; Strasser *et al.*, 2006; O'Connor *et al.*, 2007). Either of these techniques would require the collection of filter butts from survey respondents.

Purchase patterns:

Some policies influence how people obtain cigarettes. The ways in which adults change their purchase patterns after price increases, may influence the probability of subsequent quitting, with those switching to less expensive cigarettes appearing to be less likely to quit than those who do not (Hyland *et al.*, 2005; see Section 5.1 for items assessing adult purchase patterns). Among young people, policies are often enacted to reduce sales to minors (underage persons) (Lantz *et al.*, 2000). These policies are not considered effective on their own (Fichtenberg & Glantz, 2002b; Fielding *et al.*, 2005), in part because young people are more likely to give other people money to purchase cigarettes for them when restrictions on sales to minors are implemented (Everett Jones *et al.*, 2002; White & Hayman, 2006). See Table 3.14 for questionnaire items on adolescent purchase patterns.

Quit attempts

A key outcome indicator of a policy is whether it leads to an attempt to discontinue use (Starr *et al.*, 2005; Fong *et al.*, 2006a). As shown in Table 3.15, questionnaire items that assess whether a respondent has ever tried to quit, the number of lifetime quit attempts, and the duration and recency of the last quit attempt are drawn from the ITC baseline survey. ITC follow-up assessments determine whether a respondent has tried to quit since the prior assessment and the longest period of abstinence during that time period. The GATS question assesses whether a quit

Construct	Construct III.b. on Table 3.1(Type of product use)
Measure	"During the past 30 days, on how many days did you use any other form of tobacco, such as [COUNTRY SPECIFIC EXAMPLES]?" 0 days; 1 or 2 days; 3 to 5 days; 6 to 9 days; 10 to 19 days; 20 to 29 days; all 30 days (GSHS) "During the past 30 days (one month), did you use any form of smoked tobacco products other than cigarettes (e.g. cigars, water pipe, cigarillos, little cigars, pipe)?" (GYTS) "During the past 30 days (one month), did you use any form of smokeless tobacco products (e.g. chewing tobacco, snuff, dip)?" (GYTS) "Do you currently use smokeless tobacco on a daily basis, less than daily, or not at all?" (GATS) "On average, how many times a day do you use the following: [snuff by mouth, snuff by nose, chewing tobacco, betel quid, any others]?" (GATS) "In the past month, have you used any other tobacco product besides cigarettes?" IF YES: "What did you use?" FOR EACH PRODUCT USED, "How often do you currently smoke/use [PRODUCT]? Would that be daily, less than daily but at least once a week, less than weekly but at least once a month, less than monthly, or have you stopped altogether?" (ITC) "Do you currently use any smokeless tobacco such as [snuff, chewing tobacco, betel quid]?" IF YES: "Do you currently use smokeless tobacco products daily?" (STEPS – EXPANDED) "On average, how many times a day do you use [snuff by mouth, snuff by nose, chewing tobacco, betel quid, other]?" (STEPS – EXPANDED)
Source	GSHS, GYTS, GATS, ITC, STEPS
Validity	Evidence of utility. Only 2% of adolescents in Sweden who reported that they did not use cigarettes or snus during the previous month had cotinine levels ≥ 5 ng/ml (Post, 2005). It was shown that the use of cotinine and thiocyanate could distinguish smokers from smokeless tobacco users (Noland *et al.,* 1988). Kappa for use of chewing tobacco during the previous 30 days was 72.3% in CDC 14-day reliability study among high school students (Brener *et al.,* 1995).
Variation	Country-specific lists are used. In general, use of a product need not be measured in surveys if consumption of tobacco in that product is by weight < 1% of the total tobacco consumed in the country, as reported by government agricultural statistics. Exceptions to this rule can occur as, for example, when use of a particular product among youth is of concern.

GSHS: Global School Health Survey
GYTS: Global Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance
CDC: Centers for Disease Control and Prevention

Table 3.10 Type of Tobacco Product Used

Construct	Construct III.c. on Table 3.1(Brand use)
Measure	"During the past 30 days (one month), what brand of cigarettes did you usually smoke?" (SELECT ONLY ONE RESPONSE) Did not smoke cigarettes during the past 30 days; no usual brand; Add 5 most common brands; other (GYTS) "What brand did you buy when you last purchased cigarettes? Were these cigarettes filtered or non-filtered? Were these cigarettes light, mild, or low-tar?" (GATS) "Do you smoke factory-made cigarettes, roll-your-own cigarettes, or both?" IF BOTH: "For every 10 (ten) cigarettes you smoke, how many are roll-your-own? In the last month, what brand of [cigarettes/roll-your-own cigarettes] did you smoke more than any other?" [SUB-BRAND CHARACTERISTICS ARE IDENTIFIED AS NECESSARY FOR EACH NATION] (ITC)
Sources	GYTS, GATS, ITC
Validity	Face validity.
Variation	In ITC, sub-brand characteristics (e.g. length, filter versus non-filter) are identified in one of two possible ways. In many countries, such as Canada, Australia, and the United Kingdom, lists of every possible brand are developed and a code is given to each brand. The interviewer needs to determine the complete name of the brand the respondent is using. Often, the prompt, "How do you ask for your specific brand in the store?" is used to try to elicit the full name. In other countries (e.g. USA, China), where the variety of sub-brands is too great, brand names are given specific codes and interviewers determine specific sub-brand characteristics (e.g. menthol versus non-menthol, King Size, 100's, or some other length). Country-specific terms that communicate concepts similar to "light," "mild," or "low-tar" should be substituted as appropriate. These can include colour, as well as terms such as "Fine" or "Smooth." Items are adaptable for assessments of other tobacco products and for non-cigarette potential reduced exposure products (PREPs).
Comments	If necessary, country representatives should generate a list of all the brands on the market and have it available for interviewers to use to code answers. Observation of packaging to assess colour(s), presence of a legal tax stamp, and/or counterfeit brands would complement self-report.

GYTS: Global Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey

Table 3.11 Brand Characteristics

attempt of at least 24 hours was made during the previous 12 months. A baseline question from the Smoking Toolkit Study (West, 2006) assesses whether a serious quit attempt (i.e. whether the person decided to make sure they never smoked another cigarette) was ever made and, if so, the duration and recency of the last quit attempt. The follow-up questionnaires assess whether a serious attempt was made during the previous 12 months, the number of attempts, and, for up to three attempts, the recency and duration of each.

Intentionality:

Spontaneous quit attempts appeared to be more successful than those that were planned (Larabie, 2005; West & Sohal, 2006). Items assessing this construct from ITC and from the Smoking Toolkit Study (West, 2006) are presented in Table 3.16.

Dose management:

People who quit abruptly (sometimes referred to as "cold turkey")

Construct	Construct III.c. on Table 3.1(Brand Use)
Measure	"About how long have you been smoking [current brand]?" IF UNKNOWN: "Would that be less than one year, or at least one year?" (ITC) "Approximately how long have you been smoking [NAME OF CURRENT BRAND]? Before the [NAME OF CURRENT BRAND] that you smoke now, what brand did you smoke?" (AUTS)
Sources	ITC, AUTS
Validity	Face validity.
Variation	Items are adaptable for assessments of other tobacco products.
Comments	Using data from the USA, it was demonstrated that 9.2% of smokers switched cigarette brands and 6.7% switched companies during the previous year (Siegel *et al.*, 1996). Rates of switching may be higher in locations where high prices lead to smokers searching out less expensive brands. During a three year cohort study, it was observed that US adolescents who used snuff were more likely to switch from a brand with low nicotine dosage to a brand with high, than to switch from a high dosage brand to a low dosage brand (Tomar *et al.*, 1995).

AUTS: Adult Use Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey

Table 3.12 Brand Switching

appear more likely to succeed than those who gradually reduce the number of cigarettes they smoke each day (Fiore *et al.,* 1990; Gritz *et al.*, 1999). Items assessing this construct from the ITC and the Smoking Toolkit Study (West, 2006) are presented in Table 3.17.

Maintenance of abstinence versus return to use:

Discontinuing use of tobacco and maintaining abstinence are the most important disease preventing actions a user can take (US Department of Health and Human Services, 2004; Dresler *et al.,* 2006). Items assessing duration of abstinence are presented in Table 3.18.

Key constructs to measure

Several reports describe important constructs for tracking progress in reducing smoking prevalence (US Department of Health and Human Services, 1989, 1990, 1994, 1998, 2001; WHO, 1998a; Husten *et al.,* 1998; Pierce *et al.*, 1998b; Warren *et al.*, 2000; Burns *et al.*, 2000; Johnston, 2001; Kopstein, 2001;Giovino, 2002; Global Youth Tobacco Survey Collaborating Group, 2002; Godeau *et al.*, 2004; Hibell *et al.*, 2004; Global Tobacco Surveillance System Collaborating Group, 2005; Starr *et al.*, 2005; Trosclair *et al.*, 2005; Hublet *et al.*, 2006; Johnston *et al.*, 2006; Mochizuki-Kobayashi *et al.*, 2006; Warren *et al.*, 2006; White & Hayman, 2006; WHO, 2007a). Table 3.19 contains a list of key constructs to measure in prevalence surveys. The key constructs involve current use. Since current use is influenced primarily by initiation and cessation, these constructs are included as well.

Two constructs, both used in adult surveys, that are too complex to include in Table 3.19 will be presented here. GATS questions permit a six category classification of use status: 1) current daily use; 2) current non daily use – formerly daily; 3) current use - never daily; 4) former daily use; 5) former use - never daily; and 6) never used. These categories can be defined based on answers to three questions: 1) "Do you *currently* smoke [use smokeless] tobacco on a daily basis, less than daily, or not at all?;" 2) "Have you smoked [used smokeless] tobacco daily in the

Construct	Construct III.D. on Table 3.1(Intensity of use)
Measure	**Youth Surveys** "How frequently have you smoked cigarettes during the LAST 30 DAYS?" Not at all; less than 1 cigarette per week; less than 1 cigarette per day; 1-5 cigarettes per day; 6-10 cigarettes per day; 11-20 cigarettes per day; more than 20 cigarettes per day (ESPAD) "During the past 30 days (one month), on the days you smoked, how many cigarettes did you usually smoke?" I did not smoke cigarettes during the past 30 days (one month); less than 1 cigarette per day; 1 cigarette per day; 2 to 5 cigarettes per day; 6 to 10 cigarettes per day; 11 to 20 cigarettes per day; more than 20 cigarettes per day (GYTS) **Adult Surveys** "On average, how many of the following do you smoke each <day/week>?" Manufactured cigarettes; hand-rolled cigarettes; pipes full of tobacco; cigars, cheroots, cigarillos; water pipe rocks (GATS) "On average, how many cigarettes do you smoke each <day/week/month>, including factory-made cigarettes and roll-your-own cigarettes?" (ITC) "On average, how many of the following do you smoke each day?" Manufactured cigarettes; hand-rolled cigarettes; pipes full of tobacco; cigars, cheroots, cigarillos; other (STEPS)
Sources	ESPAD, GYTS, GATS, ITC, STEPS
Validity	Evidence of utility. In several countries, cotinine levels increased with increasing cigarettes per day (CPD) and levelled off between 10-20 CPD (Caraballo *et al.*, 1998; Blackford *et al.*, 2006). Indicators of nicotine dependence are associated with smoking intensity in adolescents (O'Loughlin *et al.*, 2003) and adults (Shiffman *et al.*, 2004). Kappa for smoking $\geq$ 1 cigarette/day during the previous 30 days was 76.2% in CDC 14-day reliability study among high school students (Brener *et al.*, 1995).
Variation	Items are adaptable for assessments of other tobacco products. Smokeless tobacco is measured in GATS in terms of the number of times the respondent uses a given product each day.
Comments	Intensity is the number of cigarettes/cigars/pipes full of tobacco smoked each day for daily smokers and on the days smoked for less than daily smokers (Marcus *et al.*, 1993; Centers for Disease Control and Prevention, 1994a).

ESPAD: European School Survey Project on Alcohol and Other Drugs
GYTS: Global Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance
CDC: Centers for Disease Control and Prevention

Table 3.13 Intensity of Use (Number of Cigarettes or Other Tobacco Products Smoked During a Selected Time Period)

Construct	Construct III.f. on Table 3.1(Purchase patterns)
Measure	"During the past 30 days (one month), how did you usually get your own cigarettes?" (SELECT ONLY ONE RESPONSE) I did not smoke cigarettes during the past 30 days (one month); I bought them in a store, shop or from a street vendor; I bought them from a vending machine; I gave someone else money to buy them for me; I borrowed them from someone else; I stole them; an older person gave them to me; I got them some other way (GYTS) "During the past 30 days (one month), did anyone ever refuse to sell you cigarettes because of your age?" I did not try to buy cigarettes during the past 30 days (one month); yes, someone refused to sell me cigarettes because of my age; no, my age did not keep me from buying cigarettes (GYTS) "In the area where you live, do you know of any places that sell single or loose cigarettes?" Yes; No (GYTS – OPTIONAL) "Where, or from whom, did you get the last cigarette you smoked?" Tick only one box: I didn't buy it… My parents gave it to me; my brother or sister gave it to me; I took it from home without my parent(s) permission; friends gave it to me; I got someone to buy it for me; other (specify) OR I bought it…at a hotel, pub, bar, tavern, RSL club; at a supermarket; at a news agency; at a milk bar or delicatessen; at a convenience store (e.g. Food Plus); at a tobacconist/tobacco shop; at a take-away food shop; at a petrol station; through the internet; other (specify) (ASSAD) "If you bought your last cigarette, was it from a coin-operated (vending) machine?" (ASSAD) "Sometimes people break open a packet of cigarettes and sell single cigarettes. In the last four weeks, have you bought cigarettes that were not in a full packet (for example, buying one or more cigarette(s) at a time)?" IF YES: "Thinking of the last time you bought cigarettes that were not in a full packet, where did you buy the cigarette(s) from?" I bought the cigarette(s) at a shop; I bought the cigarette(s) from a friend or relative; I bought the cigarette(s) from someone else (ASSAD)
Sources	GYTS, ASSAD (White & Hayman, 2006)
Validity	Face validity.
Variation	Items are adaptable for assessments of other tobacco products.
Comments	Those who purchase in locations that provide less expensive cigarettes are less likely to quit (Hyland *et al.*, 2005). Young people are more likely to have other people purchase cigarettes for them in regions where sales to minors are restricted (Everett Jones *et al.*, 2002; White & Hayman, 2006).

GYTS: Global Youth Tobacco Survey
ASSAD: Australian Secondary Students' Alcohol and Drug Survey

Table 3.14 Purchase Patterns

Construct	Construct IV.b. on Table 3.1 (Quit attempts)
Measure	**Ever:** ITC BASELINE: "Have you ever tried to quit smoking?" IF YES: "How many times have you ever tried to quit smoking? How long ago did your most recent serious quit attempt end? Thinking about your last serious quit attempt, how long did you stay smoke free?" (ITC) "Have you ever made a serious attempt to stop smoking? By serious attempt I mean you decided that you would try to make sure that you never smoked another cigarette." Yes; No; Don't know IF YES: "Thinking back to your most recent attempt to quit smoking, how long ago was it?" SHOW SCREEN: Within the last week; within the last 2-3 weeks; a month ago; more than 1 month and up to 2 months; more than 2 months and up to 3 months; more than 3 months and up to 6 months; more than 6 months and up to a year; more than one year and up to 5 years; longer than 5 years; don't know. AND: "How long ago did your most recent quit attempt last?" Less than a day; more than a day but less than 3 days; more than 3 days up to a week; more than a week up to a month; more than 1 month and up to 2 months; more than 2 months and up to 3 months; more than 3 months and up to 6 months; more than 6 months and up to a year; more than one year and up to 5 years; more than 5 years; don't know; I am still not smoking (STS Baseline Questionnaire) **Past 12 months:** "During the past year, have you ever tried to stop smoking cigarettes?" I have never smoked cigarettes; I did not smoke during the past year; yes; no (GYTS) "During the past 12 months, have you tried to stop smoking?" IF YES: "Thinking about the last time you tried to quit, how long did you stop smoking?" (GATS) **Follow-up assessments in a cohort study:** ITC FOLLOW-UP WAVES: FOR RESPONDENTS WHO WERE CURRENTLY SMOKING AT THE PREVIOUS WAVE: "Have you made any attempts to stop smoking since we last spoke with you in [month of last interview]?" IF YES: "Are you back smoking or are you still stopped?" IF BACK SMOKING: "What is the longest time that you stayed smoke free since [month of last interview]?" IF STILL STOPPED: "When did you quit?" (ITC) FOR RESPONDENTS WHO WERE ABSTINENT AT THE PREVIOUS WAVE: "The last time we spoke with you in [month of last interview] you had quit smoking. Are you back smoking or are you still stopped?" IF BACK SMOKING: "What is the longest time that you stayed smoke free since [month of last interview]?" IF STILL STOPPED: "So you have quit smoking since [quit date reported previously] – is that correct?" IF NO: "When did you quit?" (ITC) "Have you made a serious attempt to stop smoking in the past 12 months? By serious attempt I mean you decided that you would try to make sure that you never smoked another cigarette. Please include any attempt that you are currently making." Yes; no; don't know. IF YES: "How many serious attempts to stop smoking have you made in the last 12 months?" (Choose one option only) 1 attempt; 2 attempts; 3 attempts; more than 3 attempts; don't know. "How long ago did your quit attempt start?" (assessments are made for up to 3 attempts). "How long ago did your quit attempt last before you went back to smoking?" (assessments are made for up to 3 attempts; "still not smoking" is an option) (STS Wave 1 and 2 postal questionnaires)
Sources	ITC; STS (West, 2006); GATS

Table 3.15 Quit Attempts

Validity	Face validity. However, respondents appear to forget many short quit attempts, especially those that took place more than three months before the interview (Gilpin & Pierce, 1994; West *et al*, 2007). Having ever quit for ≥ 12 months or having quit for ≥ 7 days during the previous 12 months has been classified as a strong quitting history and is predictive of subsequent cessation (Pierce *et al*., 1998b).
Variation	Items are adaptable for assessments of other tobacco products.
Comments	ITC items are specifically crafted to assess change in a cohort study.
Definitions	A quit attempt is an activity by a user in which the person tries to stop using with the intention of never using again. Some surveys only classify periods of abstinence as quit attempts that last for ≥ 24 hours.

GYTS: Global Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
STS Smoking Toolkit Study

Table 3.15 Quit Attempts

past?;" and 3) "In the past, have you smoked [used smokeless] tobacco on a daily basis, less than daily, or not at all?" (Note that respondents are skipped past questions that do not apply to them, as indicated by their answer(s) to initial item(s).)

The second construct involves a technique that assesses tobacco use activity during the 12 months prior to being interviewed. The US Tobacco-Use Supplement to the Current Population Survey asks current daily smokers, current non-daily smokers, and former smokers abstinent ≤ 12 months, "Around this time 12 months ago were you smoking cigarettes every day, some days, or not at all?" This question, which can be adapted to smokeless tobacco use, enables a retrospective cohort assessment of cessation activity, transitioning from daily to non-daily use, transitioning from non-daily to daily use, and relapse to daily or non-daily use (Gilpin & Pierce, 1994; US Department of Health and Human Services, 1998; Burns *et al*., 2000).

Summary

This section describes the key concepts within the natural history of tobacco use, providing a conceptual model to guide measurement of key constructs. Current tobacco use is the most important construct because of its importance as an outcome in policy evaluation studies. Studies that have examined the validity of self-reported measures of current use generally find these measures to be valid, although there are conditions where the validity may be reduced.

It is important to measure the type of tobacco used, particularly in those countries in which there exists a variety of forms. The variety of forms available, and the possibility of switching, or multiple concurrent use may influence the probability of quitting and disease risk.

Detailed measurement of information about tobacco product packaging is important in order to determine the variant of product type used, movement between price sectors, and, potentially, to assess the use of tobacco from illicit sources.

Other important constructs in the measurement of tobacco use behaviour include early use, frequency and intensity of current use, quit attempts, and duration of abstinence among former smokers.

Consumers of survey data, in which tobacco use measures are included, should be aware of factors that can influence population estimates of tobacco use and take those into consideration when comparing estimates from surveys conducted within and across countries.

Construct	Construct IV.b.i on Table 3.1 (Intentionality)
Measure	"When you made your last quit attempt, when did you choose your quit day?" Chose it on the actual day when you stopped; chose it on the day before you stopped; chose it more than one day before; or actually decided to quit after having not smoked for some other reason (ITC) "Had you been seriously thinking about quitting in the days before you finally decided to stop, or was it a spur-of-the-moment decision?" I had already been seriously thinking about quitting; it was a spur-of-the-moment decision (ITC) "Which of the following statements best describes how your most recent quit attempt started?" SHOW SCREEN: I did not plan the quit attempt in advance; I just did it; I planned the quit attempt for later the same day; I planned the quit attempt the day beforehand; I planned the quit attempt a few days beforehand; I planned the quit attempt a few weeks beforehand; I planned the quit attempt a few months beforehand; none of these (other specify) (STS Baseline Questionnaire) Please circle which applies to each quit attempt. (Choose one response for each quit attempt) I planned the quit for later the same day or for a date in the future; I planned to quit as soon as I made the decision (STS Wave 1 & 2 postal questionnaires)
Sources	ITC; STS
Validity	Face validity. Unplanned quit attempts were more likely to succeed than planned attempts (Larabie, 2005; West & Sohal, 2006)
Variation	Items are adaptable for assessments of other tobacco products.

ITC: International Tobacco Control Policy Evaluation Survey
STS: Smoking Toolkit Study

Table 3.16 Quit Attempts – Intentionality

Construct	Construct IV.b.ii on Table 3.1 (Dose management)
Measure	"On your most recent quit attempt, did you stop smoking suddenly or did you gradually cut down on the number of cigarettes you smoked?" Stopped suddenly; cut down gradually (ITC) "Did you cut down gradually by delaying the first cigarette you had each day for longer and longer, or just by trying to smoke less and less?" By delaying the first cigarette of the day; by trying to smoke less and less; both (ITC) "Did you cut down the amount you smoked before trying to stop completely?" (Choose one response for each quit atempt) Cut down first; stopped without cutting down; cannot remember (STS)
Sources	ITC; STS
Validity	Face validity. Abstainers were more likely to stop without cutting down than were relapsers, who were more likely to quit using gradual reduction (Fiore *et al.*, 1990; Gritz *et al.*, 1999).
Variation	Items are adaptable for assessments of other tobacco products.

ITC: International Tobacco Control Policy Evaluation Survey
STS: Smoking Toolkit Study

Table 3.17 Quit Attempts – Dose Management

Construct	Construct IV.c. on Table 3.1 (Maintenance of abstinence)
Measure	"How long ago did you stop smoking?" I have never smoked cigarettes; I have not stopped smoking;1-3 months; 4-11 months; 1 year; 2 years; 3 years or longer (GYTS) "When was the last time you smoked a cigarette, even one or two puffs?" I have never smoked a cigarette; today; not today, but some time during the past week; not in the past week, but some time in the past month; 2-3 months ago; 4-6 months ago; 7-12 months ago; 1 to 4 years ago; 5 or more years ago (GYTS – OPTIONAL) "How long has it been since you last smoked regularly?" (GATS) ITC FOLLOW-UP WAVES: FOR RESPONDENTS WHO WERE CURRENTLY SMOKING AT THE PREVIOUS WAVE: "Have you made any attempts to stop smoking since we last spoke with you in [month of last interview]?" IF YES: "Are you back smoking or are you still stopped?" IF BACK SMOKING: "What is the longest time that you stayed smoke free since [month of last interview]?" IF STILL STOPPED: "When did you quit?" (ITC) ALTERNATIVE METHOD: "Have you made any attempts to stop smoking since we last spoke with you in [month of last interview]?" IF YES: "The last time we spoke with you in [month of last interview] you said that you smoked [daily/less than daily but at least once a week/less than once a week but at least once a month]. Do you still smoke [daily/less than daily but at least once a week/less than once a week but at least once a month]?" IF NO AND RESPONDENT SMOKED DAILY AT LAST INTERVIEW: "Are you now smoking at least once a week, or less than once a week, but at least once a month?" IF NO AND RESPONDENT SMOKED WEEKLY AT LAST INTERVIEW: "Are you now smoking daily or are you smoking less than once a week, but at least once a month?" IF NO AND RESPONDENT SMOKED MONTHLY AT LAST INTERVIEW: "Are you now smoking daily or less than daily, but at least once a week?" FOR RESPONDENTS WHO WERE ABSTINENT AT THE PREVIOUS WAVE: "The last time we spoke with you in [month of last interview] you had quit smoking. Are you back smoking or are you still stopped?" IF BACK SMOKING: "What is the longest time that you stayed smoke free since [month of last interview]?" IF STILL STOPPED: "So you have quit smoking since [quit date reported previously] – is that correct?" IF NO: "When did you quit?" (ITC) "How long ago did you stop smoking daily?" (STEPS)
Sources	GYTS, GATS, ITC, STEPS
Validity	Evidence of utility. Self-reports of having quit are reasonably valid when adequate privacy is afforded and demand for abstinence is not high (Velicer *et al.*, 1992).
Variation	Items are adaptable for assessments of other tobacco products.
Comments	ITC items are specifically crafted to assess change in a cohort study.
Definitions	A former user is someone who has used more than the threshold level of established use and who no longer uses. Sustained former use occurs when a former user has been abstinent for at least 12 months (6 to 12 months, Starr *et al.*, 2005; ≥ 12 months, Giovino & Borland, personal communication).

GYTS: Global Youth Tobacco Survey
GATS: Global Adult Tobacco Survey
ITC: International Tobacco Control Policy Evaluation Survey
STEPS: STEPwise Approach to Chronic Disease Factor Surveillance

Table 3.18 Duration of Abstinence in Former Smokers

Construct	Numerator	Denominator	Comments
Initiation of Use			
Ever use	Number of ever users	Total number of respondents	A similar construct could be assessed for ever daily use.
Early initiation	Number of ever users who tried using before a given age	Number of ever users	GYTS uses 10 years old as cut-off. A similar constuct could be measured for initiation of daily use before a given age.
Transition to established use	Number of current daily users	Number of ever users	Indicates probability of transition to and maintenance of more established use. (See Johnston, 2002 for other indicators of transition)
Discontinuance	Number of former triers	Number of ever users	A similar construct could be assessed for former experimenters.
Maintenance of Use			
Current use	Number of current users	Total number of respondents	Various measures include current smoking, current smokeless tobacco use, current tobacco use, and current use of individual products. Similar constructs could be assessed for current daily use.
Frequency of use	Number of daily users	Number of current users	An "inverse" construct would define the percentage of current users who do not use on a daily basis. Some surveys describe frequent use as use on $\geq$ 20 of the previous 30 days.
Intensity of use	Number of current users who use more than a given amount	Number of current users	Cut-offs should be standardised to permit comparisons. For example, for adult cigarette smokers, use of $\geq$ 15 cigarettes/day could serve as a measure of heavy smoking. Mean numbers can also be presented.
Brand use	Number of current users who use a given brand	Number of current users	Variants could involve descriptors of roll-your-own cigarettes, Western versus domestic brands, and sub-brand characteristics as appropriate to a given nation (e.g. "light/mild," "menthol")
Purchase location	Number of current users who purchase in a given location	Number of current users	For adults, type of venue could indicate tax avoidance strategies. For youth, source of tobacco could indicate efforts

Table 3.19 Suggested Prevalence Indicators of Tobacco Use Behaviours

Cessation of Use			
Former use among ever users	Number of former uses	Number of ever users	Often called the "quit ratio" or "prevalence of cessation" this is a crude measure of quitting (Pierce *et al.*, 1989; US Department of Health and Human Services, 1989, 1990).
Sustained abstinence	Number of former users abstinent for ≥ 6 months	Number of ever users	Relapse is less likely after being abstinent for ≥ 12 months.
Making a quit attempt	Number of current users who tried to quit during the previous 12 months plus the number of former users abstinent for ≤12 months	Number of current users plus the number of former users abstinent for ≤12 months	Making a quit attempt is a dependent variable in many policy analyses
Former use for ≥ 1 months among anyone who used during the previous 12 months and made a quit	Number of former users abstinent for 1-12 months	Number of current users who tried to quit during the previous 12 months plus the number of former users abstinent for 1-12 months	Indicates ≥ 1 month of abstinence among those who tried to quit during the previous 12 months. People abstinent for < 1 month would be not included in this anlysis (Centers for Disease Control and Prevention, 1993)

Notes: The numbers in the numerator and denominator could be either the actual number of respondents in the survey or the weighted population estimate. Also, fractions would be multiplied by 100 to obtain percentages.

Table 3.19 Suggested Prevalence Indicators of Tobacco Use Behaviours

3.2 General mediators and moderators of tobacco use behaviours

Introduction

Presented in this section are a core set of general mediator and moderator variables that should be considered when evaluating tobacco control programmes and policies. A brief description and assessment of several standard measures for assessing these constructs are provided as well. Mediators are variables situated on the causal pathway between a policy and its public health impact (i.e. variables that are affected by policies and that in turn, influence health or behavioural outcomes). For instance, motivation to quit may increase after an anti-tobacco information campaign, and motivation in turn predicts whether smokers will quit. Moderators are factors not directly affected by the specific policy under scrutiny, but that moderate the effect of that policy. For example, an information campaign may be effective among one age group while being ineffective in another (Figure 3.2). Analyzing mediators sheds light on how policies and interventions have an impact; analyzing moderators aids in understanding under what conditions and in which groups they work, or do not work. In the context of policy evaluation, nothing is as practical as a good theory that explains what to measure, how to interpret the results, what course of action to take based on these results, and what consequences can be expected from these actions. To establish a list of these mediators and moderators, the Working Group (WG) drew on relevant behaviour theories (Conner & Norman, 1996) including the Social Cognitive Theory (Bandura, 1986), the Health Belief Model (Janz & Becker, 1984), the Transtheoretical Model of Change (Prochaska *et al.*, 1992), the Protection Motivation Theory (Rogers, 1975), the Theory of Planned Behavior (Ajzen, 1991), and the Prime Theory (West & Hardy, 2006). In particular, readers are referred to the theoretical framework of the International Tobacco Control Policy Evaluation Survey (ITC), which was developed specifically for the evaluation of the WHO Framework Convention on Tobacco Control (FCTC), and within which surveys can be developed and interpreted (Fong *et al.,* 2006a; Thompson *et al.*, 2006). A comprehensive list of all the psychosocial determinants of smoking behaviour would result in a long questionnaire in the context of policy evaluation. Therefore, the WG established a short list of the variables considered to be the most relevant and useful for the evaluation of tobacco control policies and interventions in general. Researchers can complement this list by adding other relevant measures, depending on the aim and cultural context of each study, and the specific interventions under evaluation.

Guiding principles in the establishment of this list were the usefulness of each measure, its influence in the published literature, and the availability of associated validation studies (which were not always available). Some measures for which no psychometric tests of validity were available were nevertheless included because of their face validity and lack of alternative validated measures. Efficiency was also an important criterion of selection: the WG chose instruments that were both brief and informative, excluding long instruments, even if they were widely used. When several comparable scales were available, the most influential one was chosen, based on the number of citations to the original articles describing these scales (Bakkalbasi *et al.*, 2006).

The psychological determinants of tobacco use and cessation range

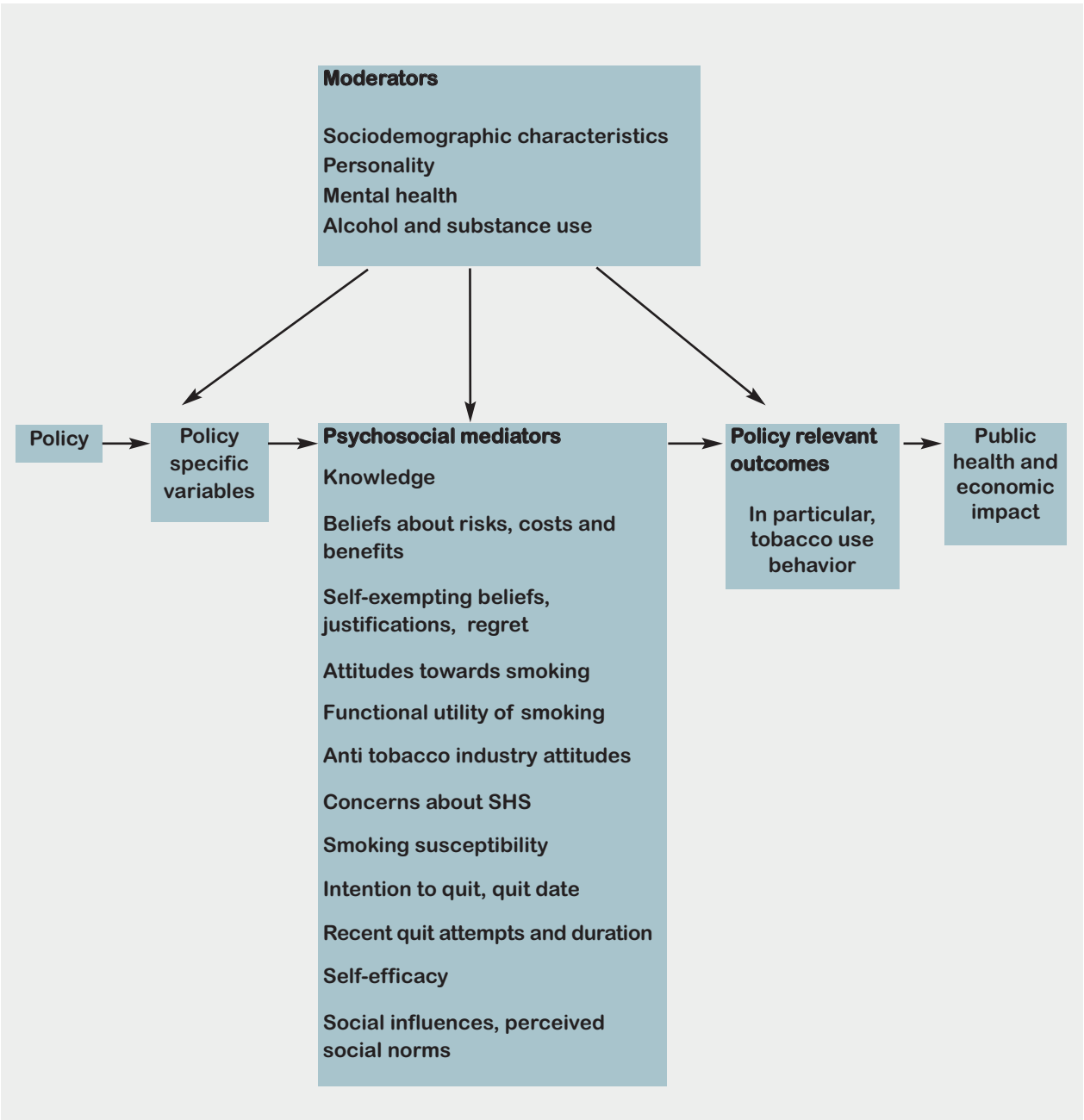


Figure 3.2 The role of psychosocial variables in the causal chain between policy and public health impact

from cognitive, motivational, and emotional variables to personality traits, personal life events, and psychopathology variables. It is important to note that many quit attempts are not planned (Larabie, 2005), that the triggers of relapse are often quite contextual, and that the timely response of the subject in each specific situation is determinant (West & Hardy, 2006). Thus, ideally, measurements should be both timely and contextual, which is not always feasible. Therefore, the WG excluded the assessment of temporary states of mind (e.g. the euphoria caused by an alcoholic drink) that are good proximal predictors of relapse, because their assessment requires specific techniques (ecological momentary assessments) that are not easily implemented in the context of policy evaluation (Shiffman *et al.,* 2002).

Smoking prevalence is much higher in psychiatric patients than in the general population, and on average, smokers with psychiatric disorders are more dependent on tobacco than other smokers (Breslau, 1995). There is also a concern that, in countries where smoking prevalence declines, an increasing proportion of the remaining smokers have psychiatric disorders (Lasser *et al.*, 2000). Thus, an assessment of mental health is relevant to the study of smoking behaviour. In addition, it is suggested that alcohol use and abuse be assessed, as both are strongly associated with tobacco use. Depending on the context, evaluators can also assess illicit drug use, for instance by using the WHO ASSIST questionnaire (WHO ASSIST Working Group, 2002; Newcombe *et al.*, 2005).

The set of general mediators and moderators considered in this section was derived from theory, published research, and the WG's subjective assessment of what is relevant for policy evaluation. This list (Table 3.20), though not comprehensive, is believed to represent a core set of measures useful in explaining how policies and interventions work, in which population subgroups they work, and how to improve them.

Items and scales used to assess the psychological determinants of smoking

Mediators

Cognitive variables

Perceived risk and outcome expectancies

For many quitters, smoking cessation is preceded by a change in beliefs about the costs and benefits of smoking and of quitting (Etter *et al.*, 2000a). These beliefs are often the target of prevention interventions, and it is therefore important to include them in programme evaluations. Assessing personalized beliefs that the respondent has about himself or herself is suggested, rather than general awareness, since personalized beliefs are stronger predictors of behaviour. Three questions are proposed to assess a respondent's perceived risk of disease: "How would you compare your chance of getting lung cancer compared to the chance of a nonsmoker?" "Do you worry that smoking will damage your health?" "How much do you think you would benefit from quitting smoking?" (Table 3.21). Additional specific beliefs are covered in other sections of this Handbook.

Validity: For the question on "worrying that smoking will damage the smoker's health," the test-retest intraclass correlation, assessed eight months apart in daily smokers with no quit attempts, was r=0.59 (Yan, 2007). In an analysis of daily smokers in the ITC surveys, this question predicted whether participants made a quit attempt (very worried versus not at all worried, odds ratio (OR) = 3.24 for quit attempts, 95% confidence interval (CI): 2.67-3.94) (Thompson *et al.*, 2006; Yan, 2007). For the question on "the benefits of quitting smoking," the test-retest intraclass correlation was r=0.54, for assessments made eight months apart in daily smokers with no quit attempts (Yan, 2007). In an analysis of daily smokers in the ITC surveys, the question on "the benefits of quitting" predicted smoking cessation after eight months (extremely versus not at all, OR = 2.11, 95% CI: 1.23-3.60) (Yan, 2007). These questions therefore have some evidence of validity.

I. Mediators

a. Cognitive variables:

- Knowledge
- Beliefs about the risks, costs, and benefits of smoking and of quitting
- Self-exempting beliefs, justifications, regret
- Attitudes towards smoking, functional utility of smoking
- Anti-tobacco industry attitudes
- Concerns about exposing others to secondhand smoke

b. Motivational variables:

- Smoking susceptibility (adolescents)
- Intention to quit and quit date
- Recent quit attempts and duration of the last quit attempt

c. Self-efficacy

d. Social influences, perceived social norms

II. Moderators

a. Sociodemographic characteristics:

- Age
- Sex
- Socioeconomic status (education, income, occupation)
- Ethnicity, primary language, minority group status
- Religion
- Family structure, peer and family smoking
- Country of residence and language of the interview (recorded by the interviewer)

b. Personality

c. Mental health:

- WHO-5 Well-Being Index
- 2-item screening for current symptoms of depression

d. Alcohol use and abuse:

- Alcohol Use Disorders Identification Test (AUDIT-C)

Table 3.20 List of Some Relevant Psychosocial Determinants of Smoking

Construct	Question and Link	Responses, Scoring	Adult / Adolescent	Recommended / Optional	Validity Level	References
	Mediators					
	Cognitive variables					
Perceived risk	Let's say that you continue to smoke the amount you do now. How would you compare your own chance of getting lung cancer in the future to the chance of a nonsmoker?	. Much more likely to get lung cancer than a nonsmoker? . Somewhat more likely . A little more likely . Just as likely	All smokers	Recommended	Face valid	ITC¯W2-175
Perceived risk	How worried are you, if at all, that smoking will damage your health in the future?	. Not at all worried . A little worried . Moderately worried . Very worried	All smokers	Recommended	Validated	ITC¯W2-180
Self-exempting beliefs	The medical evidence that smoking is harmful is exaggerated.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Optional	Validated	ITC¯W2-163
Self-exempting beliefs	You have the kind of genetic make-up that allows you to smoke without it giving you health problems.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Optional	Face valid	ITC¯W2-99
Regret	If you had to do it over again, you would start smoking again	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	Current and former smokers	Recommended	Face valid	ITC¯W2-88
Attitudes Towards smoking Scale (ATS¯18 scale)	The following are some statements concerning smoking. Please indicate whether or not you agree with each of them. 1. Smoking is extremely dangerous to my health 2. I spend too much money on cigarettes 3. My cigarette smoke bothers other people a great deal. 4. A cigarette calms me down when I am stressed. 5. After a cigarette, I am able to concentrate better. 6. I love smoking	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All smokers	Optional	Validated	Etter & Perneger (1999) Etter *et al* (2000a) Christie & Etter (2005)

Table 3.21 Measures of the Psychosocial Determinants of Smoking

Construct	Question and Link	Responses, Scoring	Adult / Adolescent	Recommended / Optional	Validity Level	References
	Mediators					
	Cognitive variables					
Benefits of quitting	How much do you think you would benefit from health and other gains if you were to quit smoking permanently in the next 6 months?	. Not at all . Slightly . Moderately . Very much . Extremely	All smokers	Optional	Validated	ITC-W2-133
Functional utility	Smoking helps you to control your weight	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Recommended	Validated	ITC-W2-156
Functional utility	Smoking calms you down when you are stressed or upset	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Optional	Validated	ITC-W2-152 Etter *et al.* (2000a)
Anti-industry attitudes	Tobacco companies can be trusted to tell the truth about the dangers of their products	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Recommended	Validated	ITC-W2-169
Anti-industry attitudes	Tobacco companies have tried to convince the public that there is little or no health risk from secondhand smoke.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Recommended	Validated	ITC-W2-172
Concerns SHS	Your cigarette smoke is dangerous to those around you.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All smokers	Recommended	Validated	ITC-W2-150
Concerns SHS	In the last month' how often, if at all, did you think about the harm your smoking might be doing to other people?	. Never . Rarely . Sometimes . Often . Very often	All smokers	Optional	Validated	ITC-W2-45d

Table 3.21 Measures of the Psychosocial Determinants of Smoking

Construct	Question and Link	Responses, Scoring	Adult / Adolescent	Recommended / Optional	Validity Level	References
	Mediators					
	Motivational variables					
Smoking susceptibility	Pierce's Smoking Susceptibility Scale (3 items) 1. (*Intention*) Do you think you will try a cigarette soon? 2. (*Efficacy*) If one of your best friends were to offer you a cigarette, would you smoke it? 3. (*Intention*) Do you think you will be smoking cigarettes 1 year from now? http://dccps.nci.nih.gov/TCRB/susceptibility.html	1: Yes, No. 2+3: . Definitely yes . Probably yes . Probably not . Definitely not	Adolescent	Recommended	Validated	Pierce *et al.* (1996)
Intention to quit	Are you seriously thinking of quitting smoking?	. No . yes, but I have not decided when . Yes, I plan to quit within the next 30 days	All smokers	Recommended	Utility	Prochaska *et al.* (1992)
Quit date	Have you set a firm date to quit smoking? If answer = Yes: When is this date?	Yes, No (mm/dd/yyyy)	All smokers	Recommended	Face valid	ITC-W5-184
Quit attempts	In the past 12 months, did you seriously try to quit smoking?	Yes, No, I don't remember	All smokers	Recommended	Face valid	Ad hoc
Duration of last quit attempt	Thinking about any quit attempt that ENDED within the last 12 months—since [12M anchor]—what is the longest time that you stayed smoke free?	Days, weeks, months, years	All smokers	Optional	Face valid	ITC-W2-44
Self-efficacy, smokers	If you decided to give up completely in the next 6 months, how sure are you that you would succeed (Responded does not need to be intending to quit to respond. Emphasize "IF" in wording)	. Not at all sure . Slightly sure . Moderately sure . Very sure . Extremely sure	All smokers	Recommended	Validated	ITC-W2-129
Self-efficacy nonsmokers	How confident are you that you will remain a non smoker	. Not at all sure . Slightly sure . Moderately sure . Very sure . Extremely sure	Nonsmokers	Recommended	Face valid	ITC-W2-129
Social influences, smokers	There are fewer places where you feel comfortable about smoking	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All smokers	Recommended	Face valid	ITC-W2-159
Social influences, former smokers	There are fewer and fewer places where you would feel comfortable about smoking	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	Former smokers	Recommended	Face valid	ITC-W2-159

Table 3.21 Measures of the Psychosocial Determinants of Smoking

Construct	Question and Link	Responses, Scoring	Adult / Adolescent	Recommended / Optional	Validity Level	References
	Mediators					
	Motivational variables					
Perceived social norms	People who are important to you believe that you should not smoke.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All smokers	Recommended	Face valid	ITC-W2-158
Perceived social norms	Society disapproves of smoking.	. Strongly agree . Agree . Neither agree nor disagree . Disagree . Strongly disagree	All	Recommended	Validated	**ITC-W2-160**
	Moderators					
Sociodemographic characteristics	Age, sex, socioeconomic status (education, income, occupation), ethnicity, primary language, minority group, religion, family structure, peer and family smoking, country of residence, language of the interview	(National census questions)	All	Recommended	Face valid	-
Smokers at home	How many people live in your household, including yourself? How many of the [number above] people currently smoke tobacco, either daily or less than every day, including yourself?	Number:__	All	Recommended	Face valid	ITC-recr-2-3
Peer and family smoking	Peer and Family smoking, 5-item scale (Pierce *et al.*, 1998) 1. Do any of your parents, step-parents, or guardians now smoke cigarettes? 2. Do you have any older brothers or sisters? 3. Do your older brothers or sisters smoke cigarettes? 4. Of your best friends who are male, how many of them smoke? 5. Of your best friends who are female, how many of them smoke? http://dccps.nci.nih.gov/TCRB/pfs.html	Answers are scored dichotomously as "no exposure" versus "exposure". No family exposure: negative response to both items 1 and 3. No exposure to friend smoking: response=zero to both items 4 and 5.	Adolescent	Recommended	Utility	Pierce *et al.* (1998c)
Friends who smoke	Of the five closest friends or acquaintances that you spend time with on a regular basis, how many of them are smokers?	Number:__	Adolescent	Recommended	Validated	ITC-W2-189

Table 3.21 Measures of the Psychosocial Determinants of Smoking

Construct	Question and Link	Responses, Scoring	Adult / Adolescent	Recommended / Optional	Validity Level	References
	Moderators					
Mental health	WHO"5 Well-Being Index Please indicate for each of the five statements which is closest to how you have been feeling over the last two weeks: Notice that higher numbers mean better well-being. *Over the last two weeks:* 1. I have felt cheerful and in good spirits. 2. I have felt calm and relaxed 3. I have felt active and vigorous 4. I woke up feeling fresh and rested 5. My daily life has been filled with things that interest me. http://www.who-5.org/	5=All of the time 4=Most of the time 3=More than half of the time 2=Less than half of the time 1=Some of the time 0=At no time	All	Recommmended	Validated	Whooley *et al.* (1997) Bonsignore *et al.* (2001) Henkel *et al.* (2003) Henkel *et al.* (2004a)
Depression screening test	2"Item screening test for current symptoms of depression: 1. During the past month' have you often been bothered by feeling down, depressed or hopeless? 2. During the past month, have you often been bothered by little interest or pleasure in doing things?	Yes, No	All	Recommended	Validated	Whooley *et al.* (1997) Kessler *et al.* (2002) Henkel *et al.* (2003) Henkel *et al.* (2004a) Henkel *et al.* (2004b)
Alcohol (AUDIT-C)	Alcobol Use Disorders Identification Test (AUDIT-C) (3 questions) http://www.oqp.med.va.gov/general/uploads/FAQ%20AUDIT-C%20for%20clinicians.doc		All	Recommended	Gold standard	Bush *et al.* (1998) Reinert & Allen (2002) Rumpf *et al.* (2002)
AUDIT-C	1. How often did you have a drink containing alcohol in the past year? Consider a "drink" to be a can or bottle of beer, a glass of wine, a wine cooler, or one cocktail or shot of hard liquor (like scotch, gin, or vodka).	Never 0 points Monthly or less 1 point 2 to 4 times a month 2 points 2 to 3 times a week 3 points 4 to 5 times a week 4 points 6 or more times a week 4 points	All	Recommended		Bush *et al.* (1998)
AUDIT-C	2. How many drinks did you have on a typical day when you were drinking in the past year?	0 drinks 0 points 1 to 2 drinks 0 points 3 to 4 drinks 1 point 5 to 6 drinks 2 points 7 to 9 drinks 3 points 10 or more drinks 4 points	All	Recommended		Bush *et al.* (1998)
AUDIT-C	3: How often did you have six or more drinks on one occasion in the past year?	Never 0 points Less than monthly 1 point Monthly 2 points Weekly 3 points Daily or almost daily 4 points	All	Recommended		Bush *et al.* (1998)

ITC W2: International Tobacco Control Policy Evaluation Survey, Wave 2 (followed by item number)
SHS: Secondhand Smoke

Table 3.21 Measures of the Psychosocial Determinants of Smoking

Self-exempting beliefs, justifications, and regret

Smokers continue to smoke, and nonsmokers start to smoke even though they are aware of the risks of smoking, in part because of self-exempting beliefs and other justifications (Chapman *et al.,* 1993; Weinstein, 1999). Quitting smoking may require shedding such beliefs and accepting information about the risks of smoking. The WG suggests including one question derived from the ITC survey, on whether people think that the medical evidence that smoking is harmful is exaggerated (Table 3.21).

Validity: In daily smokers in the ITC survey, the test-retest reliability on the question "the medical evidence... is exaggerated" was 0.64 (Yan, 2007). This question predicted smoking cessation after eight months (strongly disagree versus strongly agree, OR = 2.23, 95% CI: 1.17-4.23) (Yan, 2007). This question has some evidence of validity.

Regret

Many smokers express regret that they ever started to smoke. The WG suggests including one question on “whether the respondent would start smoking, if they had to do it over again.”

Validity: In daily smokers in the ITC survey, the test-retest correlation for this question was 0.62 (Yan, 2007). Smokers who strongly disagreed with this statement were less likely to make a quit attempt in the next eight months than those who strongly agreed (OR = 0.42, 95% CI: 0.24-0.75), but they were as likely to quit smoking (Yan, 2007). This question may nevertheless be retained because of its face validity.

Attitudes towards smoking

“Attitudes” are defined as the degree to which people have a favorable or unfavorable evaluation of smoking (Ajzen, 1991). Among the main drawbacks of smoking, as reported by smokers themselves, are the health risks, the financial costs, the bad smell, and the fact that secondhand smoke (SHS) bothers other people (Etter *et al.,* 2000a). Among the most frequently cited advantages of smoking are the pleasure to smoke, its relaxing effects, and the relief of withdrawal symptoms (Etter *et al*., 2000a). These elements are captured by several scales, for instance the Attitudes Towards Smoking Scale (ATS-18) (Etter *et al.,* 2000a); using a few items from this scale is recommended.

Validity: The ATS-18 has a robust factor structure across various samples, and test-retest correlations were high (in the range of 0.8 to 0.9) (Etter & Perneger, 1999; Etter *et al*., 2000a; Christie & Etter, 2005). The hypothesized association between attitudes and intention to quit has been reproduced in several studies (Etter & Perneger, 1999; Etter *et al*., 2000a; Christie & Etter, 2005), and a differential score (advantages minus drawbacks) prospectively predicted both smoking cessation in current smokers and relapse in former smokers, with differences between smokers and quitters ranging from 0.5 to 1.4 standard deviation units of this scale (Etter *et al*., 2000a). This scale can therefore be considered to have adequate validity (Table 3.21).

Functional utility of smoking

Many smokers use cigarettes to control their weight or as response to stress, even though tobacco withdrawal itself is a strong stressor. Two questions from the ITC survey, “whether smoking helps smokers control their weight,” and “whether smoking calms them down when they are stressed or upset,” should be included.

Validity: In a prospective sample of 272 current and former smokers, the item "smoking calms me down when I am stressed or upset" had a test-retest correlation of 0.8, and the item predicted relapse in ex-smokers (difference between abstainers and relapsers, 2.3 standard deviation units, $p<0.001$) (Etter *et al.,* 2000a). This item can therefore be considered to have adequate validity.

For the question on “whether smoking helps smokers control their weight,” the test-retest reliability (eight months apart) in smokers in the ITC survey was $r=0.74$ (Yan, 2007). In the same sample, this question predicted smoking cessation after eight

months (strongly disagree versus strongly agree, OR = 1.39, 95% CI: 1.06-1.82) (Yan, 2007). Therefore, this question has some evidence of validity.

Anti-tobacco industry attitudes

Criticism of tobacco companies is a strategy sometimes used in prevention campaigns. Good campaigns can modify attitudes towards these companies, which in turn may lower the risk of youth smoking initiation (Sly *et al.*, 2001a). Assessing anti-industry attitudes is therefore relevant in the context of programme evaluation. Two suggested items derived from the ITC surveys, "whether tobacco companies can be trusted to tell the truth about the dangers of their products", and "whether they have tried to convince the public that there is no health risk from SHS," should be included.

Validity: For the question on "whether the industry tells the truth," the test-retest reliability in smokers in the ITC survey was r=0.59 (eight months apart) (Yan, 2007). For the question on "whether the industry tried to convince the public that SHS carries no risk," the test-retest reliability was 0.45 (Yan, 2007). The figures are lower than usually recommended (Nunnally & Bernstein, 1994), but eight months may have been too long of an interval to assess test-retest for opinion items. In an analysis of daily smokers in the ITC surveys, the question on "whether the tobacco industry can be trusted to tell the truth" predicted smoking cessation after eight months (neither agree nor disagree versus strongly agree, OR = 0.65, 95% CI: 0.43-0.97). The question on "whether the industry tried to convince the public that SHS carries no risk" also predicted smoking cessation (disagree versus strongly agree, OR = 0.76, 95% CI: 0.61-0.93) (Yan, 2007). These questions have adequate evidence of validity.

Concerns about exposing others to secondhand smoke (SHS)

Decreasing exposure to secondhand smoke (SHS) is a priority of the FCTC. Policies targeting SHS may affect smokers' concerns about exposing others to it, which justifies including this topic. Two suggested questions are "whether smokers think that their smoke is dangerous to those around them," and "do smokers think about the harm their smoking might be doing to other people."

Validity: In the ITC surveys, the test-retest correlation for the item "your cigarette smoke is dangerous to those around you" assessed eight months apart in daily smokers with no quit attempts, was moderate (r=0.47) (Yan, 2007). However, in an analysis of daily smokers, this question predicted smoking cessation after eight months (strongly agree versus strongly disagree, OR = 2.59, 95% CI: 1.03-6.46) (Yan, 2007). The test-retest correlation for the item on the harm done to other people assessed eight months apart in daily smokers with no quit attempts, was also moderate (r=0.50). However, in an analysis of daily smokers, this question predicted smoking cessation after eight months (often or very often versus never, OR = 1.37, 95% CI: 1.16-1.62) (Yan, 2007). Therefore, these questions have some evidence of validity.

Motivational variables

Smoking susceptibility (adolescents)

To assess the susceptibility of taking up smoking, Pierce's Smoking Susceptibility Scale, a brief, three item, and widely cited measure intended for adolescents, is suggested (Pierce *et al.*, 1996).

Validity: Pierce's Smoking Susceptibility Scale has good predictive validity: in young never smokers, 6.5% of those with susceptibility ratings=0 had taken up smoking four years later, compared with 20.6% of those with ratings=3 (Pierce *et al.*, 1996). This scale can therefore be considered to have adequate validity, and the research papers describing it are widely cited (Pierce *et al.*, 1996; Choi *et al.,* 2001; Pierce *et al.*, 2005).

Intention to quit smoking

Intention to quit is a key predictor of smoking abstinence, as well as a key variable that policies and interventions intend to modify. Several approaches have been used to assess intention or

motivation to quit (Prochaska *et al*., 1992; Sciamanna *et al*., 2000). In particular, the concept of "stages of change" has been widely used. It proposes that people gradually progress towards smoking cessation through a series of stages, defined in particular by the level of motivation to quit (Prochaska *et al*., 1992). Indeed, the two most widely cited papers in the smoking and tobacco literature, as ranked in the report by Byrne and Chapman (2005), describe the stages of change theory (Prochaska *et al*., 1992, 1994). However, this theory has been criticized on the grounds that it does not accurately reflect reality, and that interventions based on it are no more effective than other interventions (West, 2005a). Furthermore, in the case of smokers unmotivated to quit ("precontemplators"), the stage of change theory recommends to prescribe interventions of doubtful efficacy (e.g. information on health risks) instead of effective treatments of dependence. This may be counterproductive if, for instance, the lack of motivation is due to the severity of dependence and to the intensity of withdrawal symptoms (West, 2005a). In addition, the stage of change is presented as a single variable describing behaviour change, when in fact it is a haphazard mix of four different elements (smoking status, intention to quit, past quit attempts, and duration of abstinence). Because this theory is so controversial, it should be used with caution, and reliance should instead be placed on more face valid measures of each of the four components of stages separately. Smoking status and quit attempts are discussed in Section 3.1. Intentions may fluctuate even in short intervals of time (Hughes *et al.,* 2005). Therefore, it may be preferable to ask about immediate plans to stop, since reports of plans beyond the short-term may lack validity. A single question can be used on whether smokers are seriously thinking of quitting (No; Yes, but I have not decided when; Yes, I plan to quit within the next 30 days) (Table 3.21).

Validity: In daily smokers in the ITC survey, those who were not planning to quit were much less likely to have quit eight months later than those who planned to quit in the next month (OR = 0.16, 95% CI: 0.11-0.23) (Yan, 2007).

Quit date

Setting a quit date and sticking to it is a strategy recommended to smokers in major guidelines (Fiore *et al*., 2000). A question on the planned quit date could be asked of those who plan to quit in the next 30 days (Table 3.21).

Validity: In daily smokers in the ITC survey with no quit attempts between the two assessments eight months apart, the test-retest reliability of the question on "whether smokers willing to quit had set a quit date" was low (r=0.43) (Yan, 2007). In addition, having set a quit date was not a significant predictor of cessation after eight months (no versus yes, OR = 0.75, 95% CI: 0.47-1.17) (Yan, 2007). This question can nevertheless be retained because of its face validity and usefulness, and because eight months may have been too long of an interval for analyses exploring this construct.

Previous quit attempts: Quit attempts may be affected by policy interventions, and are therefore a relevant measure for policy evaluation. Having recently made a quit attempt predicts future cessation, and the duration of the longest time off smoking is a particularly good predictor of future cessation (Ferguson *et al*., 2003; Hyland *et al*., 2006). It is worthwhile to ask smokers about the occurrence and duration of recent quit attempts.

Self-efficacy

Self-efficacy is the confidence in one's ability to stop smoking or to abstain from smoking in relapse situations (e.g. when having a drink with smokers) (Bandura, 1986). Self-efficacy predicts cessation in current smokers (Etter *et al*., 2000b) and relapse to smoking in former smokers (Gulliver *et al*., 1995). There are several multi-item scales measuring self-efficacy across various relapse situations that have satisfactory validation data, in particular, predictive validity (De Vries *et al*., 1988; Velicer *et al*., 1990; Etter *et al.,* 2000b). However, these scales are too long for the purpose of policy evaluation, and single item measures may be preferable. A single item measure of self-efficacy derived from the ITC

survey that asks "whether respondents are sure that they would succeed if they tried to quit," is suggested (Table 3.21).

Validity: The test-retest intraclass correlation for this self-efficacy item, assessed eight months apart in daily smokers with no quit attempts, was moderate (r=0.51) (Yan, 2007). However, in an analysis of daily smokers in the ITC surveys, this question predicted smoking cessation after eight months (extremely sure versus not at all sure, OR = 2.46, 95% CI: 1.68-3.59) (Yan, 2007). Therefore, this question has adequate evidence of validity.

Social influences, perceived social norms

Social influences are crucial in an adolescent's decision to take up smoking (De Vries *et al.*, 1995). In many countries, social pressures also make it less acceptable for adults to smoke (Albers *et al.*, 2004). Including three questions derived from the ITC survey to assess social influences is recommended. These questions cover "whether others who are important to the respondent believe that they should not smoke," "whether the respondent feels that there are fewer places where they feel comfortable smoking," and "the respondent's perception of the opinion that society disapproves of smoking."

Validity: The test-retest intraclass correlation for these three items, assessed eight months apart in daily smokers, was moderate (r=0.42, r=0.40, and r=0.33, respectively), but eight months may be too long of an interval to assess test-retest reliability of opinion questions. In an analysis of daily smokers in the ITC surveys, answers to the first two questions ("people believe..." and "fewer places...") were not predictive of smoking cessation after eight months (Yan, 2007). However people who agreed with "society disapproves of smoking" were more likely to have quit eight months later than people who disagreed with this affirmation (OR = 1.34, 95% CI: 1.01-1.78) (Yan, 2007). In spite of their mixed performance on validation tests, these questions can be included because of their face validity and utility.

Moderators

Socio-demographic characteristics

Sociodemographic characteristics are strong determinants of smoking behaviour (Townsend *et al.*, 1994). Relevant variables include: age, sex, marital status and social support, socioeconomic status (education, income, occupation), ethnicity, primary language, minority group status, religion, family structure, peer and family smoking, country of residence and language of the interview (recorded by interviewer).

The most appropriate questions to assess sociodemographic characteristics vary between countries (e.g. for ethnicity, minority group status, education, etc.). Using either census questions in each country or standard questions from the World Bank surveys would be recommended (Grosh & Glewwe, 1998).

Other smokers in the household, friends who smoke

Workplace and home smoking restrictions are important policy outcomes, and in turn, they are relevant determinants of smoking behaviour. The presence of other smokers in the household decreases the chances of quitting smoking (Hymowitz *et al.,* 1997), and increases the risk of smoking initiation in nonsmokers (Conrad *et al.*, 1992; O'Loughlin *et al.*, 1998; Tyas & Pederson, 1998). To assess this, it is recommended that questions about "how many people in the household are smokers," and "how many of the respondents' five best friends are smokers," be used (Table 3.21).

Validity: In the ITC survey, the test-retest intraclass correlation for the item on "how many of their five best friends smoke," assessed eight months apart in daily smokers, was r=0.64 (Yan, 2007). In an analysis of daily smokers, this question predicted smoking cessation after eight months (four friends versus 0 friends OR = 0.63, 95% CI: 0.43-0.92) (Yan, 2007). Therefore, this question has adequate evidence of validity.

Peer and family smoking (5-items), adolescents only

Peer and family smoking predicts smoking initiation in adolescents

(Conrad *et al*., 1992; O'Loughlin *et al*., 1998; Tyas & Pederson, 1998). A useful 5-item scale developed to assess the smoking status of family members and best friends has been developed (Pierce *et al*., 1998c). This widely cited scale is intended for adolescents ages 12-17, and can be administered over the phone (Table 3.21).

Validity: Peer and family smoking were not strong predictors of susceptibility to smoke (Pierce *et al*., 1998c) (OR = 1.19, non significant). Nevertheless, this scale can be used, as several other studies have shown the importance of peer and family smoking (Conrad *et al*., 1992; O'Loughlin *et al*., 1998; Tyas & Pederson, 1998). Also because this scale is widely used (cited by at least 227 articles), it enables comparison between samples.

Personality

Personality traits affect smoking behaviour. For instance, a heritable tendency for sensation seeking or for novelty seeking predicts smoking behaviour (Zuckerman *et al.,* 1990; Pomerleau *et al*., 1992; Etter *et al.,* 2003a). Most personality questionnaires are too long to be used in policy evaluation surveys (Cloninger *et al.,* 1993; Barrett *et al*., 1998); however, depending on the research goals, short versions of some personality questionnaires, such as for sensation seeking, have been validated and could be considered for inclusion (Hoyle *et al*., 2002; Stephenson *et al*., 2003).

Mental health

Smoking behaviour is strongly associated with mental health, including depression (Glassman *et al*., 1990), which justifies the inclusion of a brief assessment of mental health in surveys of the general population. Among brief assessments suitable for general population surveys, evaluators can choose, according to their specific needs, between the WHO-5 Well-Being Index, which is a measure of mental well-being (Bonsignore *et al*., 2001), and a 2-item screening test for depression (Whooley *et al*., 1997). Mental health patients are often hard to reach and may not take part in population surveys. Because particular attention should be paid to this group, population surveys should be supplemented with specific surveys of mental health patients.

WHO-5 Well-Being Index (WHO-5)

Being a WHO product, the 5-item WHO-5 Well-Being Index (WHO-5) enables its users to compare their results with other WHO surveys (Table 3.21) (Bonsignore *et al.,* 2001).

Validity: Using the Composite International Diagnostic Interview (CIDI) as the measure, WHO-5 had a sensitivity of 93% and a specificity of 64% to detect depression in primary care patients (Henkel *et al*., 2003). WHO-5 performed better than a clinical diagnosis to detect depression, using CIDI as the criterion (Henkel *et al*., 2004a), and can therefore be considered to have adequate validity.

A 2-item screening test for depression

A second way to assess depression in population surveys is to use a brief screening test, for instance, a widely cited 2-item test (Whooley *et al*., 1997). This test screens specifically for depression, whereas WHO-5 monitors a broader index of mental health. Another possibility is to use Kessler's K-6 scale (a 6-item measure of psychological distress) (Kessler *et al*., 2002). Finally, a question on whether the respondent has ever been diagnosed or treated for depression could also be included.

Validity: In patients without substance abuse, Whooley's 2-item test had a sensitivity of 96%, a specificity of 66%, and an area under the Receiver Operating Characteristic (ROC) curve of 0.84, using the Diagnostic Interview Schedule (DIS-II-R) as the criterion (Whooley *et al*., 1997). The sensitivity of this 2-item scale was better than for the Center for Epidemiologic Studies-Depression scale (CES-D short) (84%) and for the Beck Depression Inventory (BDI short) (87%), and its specificity was similar or somewhat lower (CES-D short=75%, BDI short=67%) (Whooley *et al*., 1997). In another study conducted in primary care patients, this 2 item test had a similar area under the ROC curve (0.859) compared with WHO-5 (0.862), and a comparable sensi-

tivity (92% versus 93% for WHO-5) and specificity (59% versus 64% for WHO-5), using CIDI as the criterion (Henkel *et al.*, 2004b). Whooley's 2-item screening test can therefore be considered to have adequate validity.

Alcohol use and abuse: Alcohol Use Disorders Identification Test (AUDIT-C)

Alcohol use and abuse is strongly associated with tobacco use, and, in former smokers, with relapse (Hymowitz *et al.*, 1991). This justifies the inclusion of a well-validated and widely cited test of alcohol use and abuse: the 3-item Alcohol Use Disorders Identification Test (AUDIT C) (Table 3.21) (Bush *et al.*, 1998; Reinert & Allen, 2002; Rumpf *et al.*, 2002).

Validity: The brief, 3-item version (AUDIT-C) performs as well as the full version of AUDIT to detect at-risk drinkers (Bush *et al.*, 1998; Reinert & Allen, 2002; Rumpf *et al.*, 2002). AUDIT-C has good sensitivity (54% to 98%) and specificity (57% to 93%) for various definitions of heavy drinking. AUDIT-C can therefore be considered to have adequate validity.

Discussion

An assessment of the psychosocial determinants of smoking is essential to understand how policies and interventions produce their effects, and how to improve them. Evaluation studies that neglect these elements loose an opportunity to help the field progress towards more effective and acceptable interventions. Importantly, analyzing psychosocial factors is also an issue of social inequalities. Some interventions may have adverse effects in a number of subgroups, and interventions targeted at the general population may not reach several subgroups in which smoking prevalence is particularly high (e.g. mental health patients, some minorities).

The issue of translation and cultural adaptation of the measures described in this section are addressed elsewhere in this Handbook (Section 2.2). Depending on the construct under scrutiny, even well-translated questions may not be relevant, or may not be understood in a culture distant from where the instrument was initially developed (Beaton *et al.*, 2000). Many of the measures discussed here were developed in high-income, English-speaking countries, and there are very few data on their relevance or psychometric properties in other cultures.

Establishing a list of the psychosocial determinants of smoking is an impractical task that inevitably results in a list that is too long for some purposes, and too short for others. Such a list is potentially endless. The WG selected a core set of measures with general relevance for the evaluation of tobacco control programmes and policies. Their choice was based on influential theories of behaviour change, and in particular on a model derived from these theories: the conceptual framework of the ITC project (Fong *et al.*, 2006a; Thompson *et al.*, 2006). This model was developed specifically for the evaluation of the FCTC, and it is therefore relevant for the purpose of this Handbook. The WG also included some elements believed to be important, such as mental health and substance use. Whenever possible, validated measures were included (psychometric validation studies were not always available). Some measures that were not well validated were nevertheless included because of their usefulness and face validity. The WG's selection was also based on a subjective assessment of what is useful and important. Thus, this list should be supplemented by other elements according to the specific needs of each study and country, and take into account new contributions to theory (West & Hardy, 2006). Even though this list is not comprehensive, the WG believes that it represents a core set of measures that are useful in analyzing how policies and interventions work, in which population groups they work, and why some interventions do not work. Progress in this field is possible only if thorough evaluations enlighten the path.

Summary and recommendations

This section describes mediators and moderators theorized to be important in understanding how policies and interventions affect tobacco use behaviours, and under what circumstances they

have an impact. A core set of measures likely to be important has been identified. Researchers should select from this list and, when appropriate, supplement it with other relevant measures, depending on the specific context and goals of each study. There are validated measures of many of the reviewed constructs, and researchers should, whenever possible, use them rather than develop their own ad hoc measures. Investigators should report the psychometric properties of their measurement instruments, and at least the test-retest reliability, convergent validity, and/or predictive validity. Psychological measures are particularly sensitive to wording and to cultural context; therefore, the methods for translations and cultural adaptations described in Section 2.2 should be utilised in populations where these measures have not been previously validated.

3.3 Measurement of nicotine dependence

Introduction

In this section, evidence of the validity of self-report measures of nicotine/tobacco dependence in adults is examined. Measures are concentrated on that are potentially appropriate for population-based/ epidemiologic research, as nicotine dependence is often assessed as a potential moderator of programme and policy effects. The Working Group (WG) has focused mainly on scales measuring cigarette dependence, as cigarette smoking accounts for most of the health damage caused by tobacco, and because the most widely used and best studied scales measure cigarette dependence. This section has not attempted to review evidence evaluating measures to assess nicotine dependence of other types of smoked tobacco products (e.g. cigars, pipe tobacco, bidis, hookah), although adaptations of measures used to assess cigarette smoking dependence would be reasonable to consider. The WG did include a review of measures of dependence on smokeless tobacco products, since the pattern of compulsive use of these products is similar to that observed for cigarette smoking (IARC, 2007b). Persistent use of nicotine medications has been described, but it is very rare (Shiffman *et al.*, 2003). Also, long-term use of nicotine medications has no documented untoward health effects, so therefore measurement of dependence to nicotine medications will not be included in this review. Finally, while dependence on tobacco products is clearly evident among some youth, research on measures of nicotine dependence in adolescents is limited, and will not be considered in this section. For those interested in a measure of nicotine dependence among youth, please refer to the paper which describes the measurement properties of the Hooked on Nicotine Checklist (DiFranza *et al.*, 2002b).

Nicotine dependence is a hypothetical construct that is designed to explain and predict societally-important outcomes, such as an inability to quit smoking, heavy use, and other problems occasioned by smoking or tobacco use (Piper *et al.*, 2006). Assessing tobacco dependence is difficult and is made even more so in population-based epidemiologic research by the need for efficient assessment (valid and brief). Ideally, a measure should reflect the nature or domain of the construct of interest (i.e. tobacco dependence), predict important outcomes (e.g. likelihood of quitting, problems encountered through use), and be relatively brief to assess.

Measures of cigarette-induced nicotine dependence

The following section provides a brief review of data on the measurement properties of seven self-report measures developed to assess the construct of cigarette-induced nicotine dependence: 1) Fagerström Test for Nicotine Dependence (FTND); 2) Heaviness of Smoking Index (HSI); 3) Diagnostic and Statistical Manual-IV (DSM-IV) criterion of dependence; 4) International Statistical Classification and Related Health Problems-10 (ICD-10) criteria; 5) Cigarette Dependence Scale (CDS); 6) Nicotine Dependence Syndrome Scale (NDSS); and 7) Wisconsin Inventory of Smoking Dependence Motives (WISDM).

Each measure will be evaluated based on a review of the items that constitute the scales in terms of their reading level, face validity, coverage of the dependence domain, and cross-cultural applicability. The WG will review the psychometrics of each scale, including its reliability (e.g. internal consistency) and factor structure, and will examine the predictive validity of each measure, focusing on two specific tobacco dependence criteria: a pattern of pervasive and heavy smoking and the ability to quit smoking.

Pervasive and heavy smoking could be assessed using self-report measures (e.g. cigarettes smoked per day or lifetime cigarettes smoked), or using biomarkers of exposure (e.g. carbon monoxide (CO), cotinine, puff topography) (see Section 3.1), and the ability to quit smoking could be assessed using a number of strategies as well (see Section 3.1). These criteria reflect the sheer volume of tobacco products consumed and the intransigence of drug use, both of which have significant effects on the health and economics of both the individual and society. Although it is not a validation criterion, the evidence of genetic linkages to the various measures of tobacco dependence will be examined. This information may be helpful for researchers who are interested in using epidemiological measures to make inferences regarding etiology.

It is important to note that other criteria could be used to evaluate the performance of dependence measures. For instance, such measures could be evaluated with respect to prediction of withdrawal severity or other outcomes theoretically linked to dependence (Piper *et al.*, 2006). However, such outcomes seem less relevant than the ones selected for measures to be used in epidemiologic research. For the purposes of epidemiologic research, a measure should reflect or predict outcomes of societal import, such as degree of tobacco exposure and use, the intransigence of use, and the likelihood of important negative outcomes of use. Obviously, a pattern of heavy, pervasive smoking will capture the degree of exposure to nicotine and the harmful constituents of tobacco/cigarettes. Moreover, a relative inability to quit smoking will forecast the likely continued exposure to such elements. Evidence shows that past, current, and future use of tobacco directly predict outcomes of societal import, such as money expended in buying tobacco products and disease outcomes (and associated costs) caused by tobacco use (US Department of Health and Human Services, 2004; Centers for Disease Control and Prevention, 2005).

Overarching issues:

It is important to note that dependence is a construct (i.e. a hypothetical entity). It is not, in theory, equivalent to any single measure or criterion (Piper *et al.,* 2006); although single items can be used to estimate a person's standing on the construct. Thus, dependence is an inferred influence or force that produces the outcomes associated with it (e.g. high rates of smoking, relapse), although it is not the only predictor of such outcomes. Generally it takes multiple variables or items to adequately assess a complex, hypothetical entity such as nicotine dependence (Clark & Watson, 1995). In this section, however, considerable attention is devoted to very brief measures of dependence, as evidence shows that such measures (i.e. number of cigarettes smoked per day) can predict outcomes, such as relapse, as well as longer measures (e.g. DSM-III-R, FTQ, and FTND) (Razavi *et al.,* 1999; Breslau & Johnson, 2000; Dale, *et al.*, 2001).

When considering the information comprised here, it is important to remember that reliability and validity are not inherent in measures. It can not be assumed that one can generalize psychometric properties across different use contexts, or that validity for one use of a measure is generalizeable for a different use (e.g. predicting relapse likelihood versus withdrawal severity). Rather, these features are estimated based on patterns of statistical covariation and are influenced by the nature of the population being assessed (Nunnally & Bernstein, 1994; McDonald, 1999). For instance, there may be less variance in item scores, or item scores might have a less skewed distribution, when a dependence measure is used in a clinical population rather than a nationally representative population. This could easily affect both reliability and validity estimates. Different populations might yield different psychometric data because of true differences in the severity or range of dependence. However, differences might also arise because of other factors, such as secular or environmental events that might affect scores on dependence measures, while not actually changing the dependence per se. One study showed that US smokers had higher frequencies of severe nicotine dependence

(FTND ≥ 6) than did Spanish smokers (de Leon *et al*., 2002). It is possible that such population differences reflect different degrees or sources of error across the two populations (restrictions in smoking in the home, the amount of discretionary income, gender differences in smoking across the populations, the ways the smokers answer the questions and, indeed, understand them and so on) rather than differences in the biological/psychological internal processes that make up dependence. There are numerous environmental or social sources of error variance that could differentially affect the validity of a measure across populations: smoking policies in the workplace, taxes, religious or social norms, to list few.

In recognition of the dependence of psychometric properties on the population being assessed, reliability and validity data from both clinical trials and epidemiologic studies conducted around the world, and present data relating to the heritability of dependence as it is assessed using the different measures, will be presented. The tobacco dependence measures will be divided into two groups: *unidimensional* and *multidimensional*. Unidimensional measures are intended to assess dependence as a single dimension (although some, it turns out, may actually be multifactorial). Such measures are useful, because the best of them are fairly efficient in that they possess significant validity given their length/response burden. In fact, as efficient as some of the unidimensional measures are, some data suggest that particular items from these measures possess predictive validities that meet or exceed those of the whole measure (Storr *et al.,* 2005). Such items might be especially valuable for epidemiologic research.

A review of multidimensional measures of nicotine dependence are included despite their length and reduced efficiency, because they have the potential to provide information about the mechanism underlying nicotine dependence not supplied by unidimensional measures. For instance, multidimensional measures are intended to assess particular facets of dependence or dependence processes (e.g. particular motives for drug use). Thus, these measures may provide greater insights into the nature of tobacco dependence than do unidimensional measures. They also may provide greater discrimination amongst smokers/tobacco users to the extent that smokers may be distinguished on the basis of something other than a single intensity dimension (which might be well captured by a single severity dimension). For instance, some scales appear to reflect motives associated with initial versus extensive use of tobacco (Piper *et al*., 2004), and other scales differ in sensitivity to use patterns of highly dependent users versus “chippers” (those who engage in periodic or light tobacco use) (Shiffman & Sayette, 2005). Since the subscales of multidimensional measures tend to ask about relatively discrete processes (e.g. a taste motive for smoking) rather than global consequences of smoking (e.g. smoking causing problems in life), these multidimensional measures may be more suitable for genetics research, as they may tap processes that reflect a stronger genetic signal (Baker *et al*., in press). Finally, because multidimensional measures tend to ask about internal and subjective phenomena (e.g. role of affect regulation) rather than externally referenced events (e.g. latency to smoke in the morning, number of cigarettes consumed each day), these measures may be less susceptible to biasing by error due to regional secular or policy influences. Workplace smoking restrictions, for example, might exert a more direct and larger effect on number of cigarettes smoked per day than on the smokers liking of the taste of cigarettes. On the other hand, multidimensional scales tend to ask about relatively subtle, psychological variables (e.g. asking individuals to attribute smoking urges or affect), and it is possible, indeed probable, that cultures may differ in how they make attributions or label internal phenomena. Of course, while entire multidimensional scales can be quite lengthy, individual items or subscales can be selected for use (Lerman *et al*., 2006); thus, this section will review relevant subscale data.

The foregoing discussion should make clear that blanket

recommendations cannot be given regarding dependence. Rather, the investigator must both weigh practical issues (e.g. response burden) and clearly identify the goals of assessment (e.g. predict probability of relapse) in order to select an appropriate dependence instrument or assessment strategy.

Unidimensional measures of tobacco dependence

Fagerström Test for Nicotine Dependence and the Heaviness of Smoking Index

The first unidimensional measure of tobacco dependence is actually a group of measures arising from the Fagerström Tolerance Questionnaire (FTQ) (Fagerström, 1978): these comprise the FTQ itself, as well as the 6-item Fagerström Test for Nicotine Dependence (FTND) (Heatherton *et al*., 1991) and the 2-item Heaviness of Smoking Index (HSI) (Kozlowski *et al*., 1994). See Appendix 1 for the items and scoring. These measures are based on the construct of physical dependence, which includes facets such as the need to smoke early in the morning to alleviate overnight withdrawal, the need to smoke numerous cigarettes per day, and the invariance of smoking behaviour (i.e. smoking even when you are ill) (Fagerström, 1978). The Flesch-Kincaid Reading Grade Level is 4.4 for the FTND and 4.2 for the HSI.

The FTND has been translated and used with population samples in Germany (John *et al*., 2003a; John *et al*., 2004a), Switzerland (Etter *et al.,* 1999), Australia (Pergadia *et al*., 2006a), Canada (Howard *et al*., 2003), Austria (Lesch *et al.,* 2004), and Brazil, Mexico, Poland, and China (Blackford *et al*., 2006; Huang *et al*., 2006). The HSI has also been used in research in Spain (Diaz *et al*., 2005), Australia, Canada, UK, and the USA (Heatherton *et al*., 1991; Hymowitz *et al*., 1997; Hyland *et al*., 2006). One of the questions on the FTND concerns smoking in forbidden places. The validity of this question may be affected by regional differences in environmental restrictions in smoking (Huang *et al*., 2006). In addition, two questions in this scale assume a pattern of daily smoking (e.g. questions 1 & 4, the two questions in the HSI). It is very likely that scores on these items will have reduced validity if used with non-daily smokers. An important goal of future research is to identify dependence measures that are appropriate for non-daily smokers.

Reliability and structure: Compared with the FTQ, the FTND has demonstrated better psychometric properties, such as internal consistency (Payne *et al.,* 1994; Pomerleau *et al*., 1994; Haddock *et al*., 1999); however, these improved reliability coefficients are still low (Etter, 2005) and below traditionally accepted standards for clinical use (α =0.80) (Nunnally & Bernstein, 1994). Using a French translation of the FTND with light smokers found internal consistencies of approximately 0.70 (Etter *et al*, 1999), while a study with a German population found low internal consistency for the FTND (α =.57) in two separate samples (John *et al*., 2004b), and a study in China found that FTQ had low internal consistency as well (α =.58) (Huang *et al*., 2006).

Some studies have shown that the FTND has a two-factor structure, suggesting that it does not measure a unitary construct of physical dependence (Payne *et al*., 1994; Etter *et al*., 1999; Haddock *et al*., 1999; Radzius *et al.,* 2003; Breteler *et al*., 2004; John *et al*., 2004b). A population-based study in France found that while a two-factor model fit the data well, the two factors were highly correlated (Chabrol *et al*., 2003). Inter-item correlations also reveal that not all items are highly related (r = 0.06-0.39) (Transdisciplinary Tobacco Use Research Center (TTURC) Tobacco Dependence Phenotype Workgroup, 2007). These studies suggest that the two factors reflect *morning smoking* (i.e. whether one smokes more in the morning and whether one would rather give up the first cigarette of the day or all others), and *smoking pattern* (i.e. the number of cigarettes smoked per day, time to first cigarette, difficulty refraining from smoking, and smoking when ill), although some data indicate that time to first cigarette loaded on both factors (Radzius *et al*., 2003). Latent class analyses suggest that the FTND divides smokers into groups based on severity of dependence (Storr *et al*., 2005); that is the two factors do not

appear to "pick-out" smokers who differ in terms of types of dependence.

The HSI is comprised of only two items, which limits the relevance of internal consistency estimates. However, zero-order correlations between the two items in the measure indicate moderate levels of association (e.g. r's ≈ 0.30) (TTURC Tobacco Dependence Phenotype Workgroup, 2007).

Validation: The FTND and HSI predict both behavioural and biochemical indices of smoking in Chinese-, English-, French-, and German-speaking populations (e.g. CO, cotinine, lifetime amount smoked) (Heatherton *et al*., 1989, 1991; Kozlowski *et al*., 1994; Etter *et al*., 1999; John *et al*., 2003a; Huang *et al*., 2006). This should not be surprising, given that the FTND and HSI directly assess smoking heaviness. However, it is encouraging to note that smokers are indeed able to estimate their amount of smoking as indexed by biochemical tests in response to single items (e.g. Question #4 on the FTND, "How many cigarettes/day do you smoke?"). The FTND has demonstrated an ability to predict cessation outcomes in smoking cessation studies (Campbell *et al.,* 1996; Westman *et al.,* 1997; Alterman *et al.,* 1999; Patten *et al*., 2001; TTURC Tobacco Dependence Phenotype Workgroup, 2007), and with college students in a population-based study (Sledjeski *et al*., 2007). In addition, the FTND has been shown to index a heightened risk for psychiatric comorbidities in a large population sample in Germany (John *et al*., 2005).

Some data indicate that the standard scoring method used with the FTND (adding up item responses) may not produce an optimal scaling of dependence level. Latent class analysis suggested that some items are particularly important to the assessment of dependence level (those that capture variance due to morning smoking) and that they are relatively underweighted in the typical scoring method (Storr *et al*., 2005). Therefore, investigators using the FTND may wish to explore alternative, empirically-based scoring or cut-score determination methods (e.g. latent class analysis, Receiver Operating Characteristic curves (Swets *et al*., 2000)).

While the FTND certainly can predict future smoking or likelihood of cessation, the HSI appears to account for much of the predictive validity of that measure (Breslau & Johnson, 2000; Heatherton *et al*., 1989; TTURC Tobacco Dependence Phenotype Workgroup, 2007). Population-based studies conducted in Australia, Canada, the UK, and the USA found that the two HSI items (number of cigarettes smoked and time to first cigarette in the morning) were the strongest predictors of quitting (Hymowitz *et al*., 1997; Hyland *et al*., 2006). Furthermore, recent research has shown that a single item on the FTND and HSI (Item #1 – latency to first cigarette in the morning) predicts relapse vulnerability, as well as, or better than, much longer multidimensional instruments (TTURC Tobacco Dependence Phenotype Workgroup, 2007). Recent population-based research shows that a single item on the HSI (item #1) is highly effective in predicting the likelihood of future cessation (TTURC Tobacco Dependence Phenotype Workgroup, 2007).

Heritability: In a study of young adult Australian Twins, HSI-assessed dependence was found to be highly heritable (71%) (Lessov *et al*., 2004). In addition, the FTND and HSI were both related to the dopa decarboxylase gene, which is involved in the synthesis of dopamine, norepinephrine, and serotonin (Ma *et al*., 2005). One haplotype was significantly related to dependence in both African-American and Euro-American smokers, while another was related to dependence only in Euro-American smokers (Ma *et al*., 2005). Additional studies link FTND-defined dependence to particular genetic variants (Bierut *et al*., 2007; Gelernter *et al*., 2007; Saccone *et al*., 2007).

Summary: The FTND has been widely used in a number of different countries and a number of different languages. It is short and has an accessible reading level. In addition, while there are concerns regarding its structure and reliability, it has been found to predict smoking heaviness and cessation outcome. However, it

appears that the HSI is a more efficient predictor of outcome than is the FTND (using only two items). FTND and HSI scores have also been found to be heritable and related to specific dependence-linked genetic variants.

The Diagnostic and Statistical Manual, International Statistical Classification of Diseases and Related Health Problems, 10th Revision and the Tobacco Dependence Screener

Two different diagnostic systems are commonly used to diagnose tobacco dependence: both are typically considered to be unidimensional measures of tobacco dependence. One is the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) (American Psychiatric Association, 1995)[1] which is based on an empirically driven, syndromal medical model, rather than on a theoretical model of dependence (see Appendix 2 for the criteria). The second is the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), an international diagnostic classification system that was endorsed by the 43rd World Health Assembly in May 1990 and came into use by WHO Member States as of 1994 (see Appendix 3 for the criteria (WHO, 1993)). The Tobacco Dependence Screener (TDS) (Kawakami *et al.*, 1999) is a 10-item, self-report questionnaire designed to assess ICD-10, DSM-III-R (the 1987 revision of DSM-II), and DSM-IV symptoms of dependence with a Flesch-Kincaid Reading Grade Level of 8.1 (see Appendix 4 for items and scoring). To the best of our knowledge, this is the only published, self-report DSM/ICD questionnaire of tobacco/nicotine dependence. Most of the existing research has utilised the DSM criteria, and that will be the focus of this Handbook's review of diagnostic classifications of tobacco dependence.

DSM and ICD structured clinical interviews, such as the World Mental Health Survey Initiative version of the Composite International Diagnostic Interview (CIDI), or the National Institute of Mental Health Diagnostic Interview Schedule (DIS), have been translated into various languages and used in at least 11 population-based studies (Hughes *et al.*, 2006) in countries including: Germany (John *et al.,* 2003b (DSM); John *et al*., 2004a (DSM); Hoch *et al.*, 2004 (DSM)), Australia (Pergadia *et al.*, 2006b (DSM)), Canada and Taipei (Howard *et al.*, 2003 (DSM)), Spain (de Leon *et al.*, 2002 (DSM)), Austria (Lesch *et al.,* 2004 (DSM & ICD)), Switzerland (Angst *et al.*, 2005 (DSM)), Japan (Yoshimura, 2000 (ICD)), Korea (Lee *et al.*, 1990 (DSM)), and the USA (Breslau *et al.*, 2004 (DSM); Hughes *et al.*, 2004a (DSM & ICD)). The ICD-10 criteria are available in 42 languages, including Arabic, Chinese, English, French, Russian, and Spanish. The DIS, CIDI, and other diagnostic interviews comprise a series of branching questions that are aimed at eliciting information about features relevant to nicotine dependence.

Some aspects of the DSM-derived interviews and similar instruments may cause problems in any sample, or when using the instrument with culturally diverse populations. Another important caveat to observe, in regards to the DSM measure of dependence, is that the scoring algorithm used in establishing formal DSM diagnoses does not appear to yield decision rules that agree with empirical methods, such as latent class analysis (Muthen & Asparouhov, 2006). Thus, the investigator may wish to explore different methods for item-weighting and cut-score estimation if a categorical outcome is desired. In addition, it should be noted that the tobacco sections of DIS and CIDI are quite long (over 30 items), and were designed to be administered either in a face-to-face interview or by a trained professional. New technology has made it possible to have individuals respond to text-based presentations of the questions, but it is unknown how valid this presentation method would be and it would remain quite time consuming.

[1]There has been a text revision of the DSM-IV (American Psychiatric Association, 2000), however this revision did not alter any diagnostic criteria for any diagnostic categories, including the substance dependence diagnosis

Reliability and structure: Data on the reliability and structure of diagnostic interview measures of nicotine dependence arise from studies using face-to-face administration strategies. Therefore, the following conclusions cannot be generalized to a different administration format. There is evidence that the various structured diagnostic measures yield reliable diagnoses as assessed by test-retest reliability (κ = 0.63, Grant *et al.*, 2004; κ = 0.88, Hughes *et al.*, 2004a; κ = 0.73, Koenen *et al.*, 2005). One-factor analysis indicated that responses to the CIDI had a strong single factor structure (Strong *et al.*, 2003); although other factor analyses of the structured diagnostic items found that a two-factor structure was a better fit (Johnson *et al.*, 1996; Radzius *et al.*, 2004; Muthen & Asparouhov, 2006). Patterns of covariation that were found amongst the symptoms could be best accounted for by two factors (Muthen *et al.*, 2006). The first accounted for covariance in the "tolerance," "larger amounts," and "time spent using" items (see Appendix 2). Thus, this factor seems to be highly related to sheer amount smoked. The second factor was related to "persistent desired/unsuccessful efforts to cut down or quit," and "continued use despite emotional/-physical problems." Confidence in this solution is bolstered by the fact that it was obtained in three separate groups of individuals. It is also consistent with other recent factor analyses (Lessov *et al.*, 2004). Investigators might wish to analyze these item parcels separately since they may be addressing somewhat distinct constructs.

The TDS, a written questionnaire assessing the presence of diagnostic criteria, has demonstrated acceptable internal consistency in Japanese smokers (α = 0.74-0.81) (Kawakami *et al.*, 1999), but was less internally consistent among smokers in the USA (α = 0.64) (Piper *et al.*, 2008). To date, there have been no studies comparing the reliability of the interview measures with the paper-pencil measure. Therefore, one cannot assume that the psychometric data generated by the interview-format delivery of DSM or ICD items would generalize to a self-administered format.

Validation: Evidence suggests that the small set of dichotomous DSM items can distinguish between light versus heavy smoking (Strong *et al.*, 2003). An epidemiological study found that the DSM-III-R (as assessed by the DIS), was a significant, though weak, predictor of cigarette abstinence over one year, but that the FTND was a better predictor and that number of cigarettes smoked per day was the best predictor (Breslau & Johnson, 2000). Another study showed that DSM-IV diagnoses of nicotine dependence predicted heaviness of use and cessation outcome in a population-based study of college students (Sledjeski *et al.*, 2007). Several studies have shown that DSM-IV nicotine dependence diagnosis is associated with greater risk of psychiatric comorbidities in adults and youth (Grant *et al.*, 2004; John *et al.,* 2004a; Dierker *et al.*, 2006). In addition, DSM diagnoses of nicotine dependence were significantly associated with self-rated general health in a population sample in Germany (John *et al.*, 2005). In sum, there is substantial evidence that DSM/ICD diagnoses are meaningfully related to smoking heaviness and a variety of health outcomes.

Studies have shown that the TDS is associated with the smoking heaviness measures (e.g. number of cigarettes smoked per day, CO levels) and years of smoking (Kawakami *et al.*, 1999; Piper *et al.*, 2004). With respect to relapse, one study found that Japanese smokers with lower TDS scores were more likely to quit smoking after a health risk appraisal (Kawakami *et al.*, 1999). However, data from smokers who participated in smoking cessation studies in the USA, revealed that the TDS did not predict abstinence at 1-week or 6-months post-quit (TTURC Tobacco Dependence Phenotype Workgroup, 2007).

Heritability: There has been considerable research supporting the heritability of DSM/ICD-diagnosed nicotine dependence. In the Australian Twin sample study, analyses revealed that all of the DSM-IV symptoms and diagnosed DSM-IV dependence were meaningfully heritable (45-73%), and that the DSM-IV criteria of

tolerance, withdrawal, and difficulty quitting were the most highly heritable symptoms of nicotine dependence for both men and women (Lessov *et al*., 2004). A study of twin fathers, using the Vietnam Era Twin Registry, found that paternal DSM diagnosis of nicotine dependence was significantly associated with offspring DSM diagnosis of nicotine dependence (Volk *et al.,* 2007). However, one study found that DSM nicotine dependence was not related to familial liability to smoking persistence, because familial density of persistence was not associated with smoking persistence among nicotine-dependent daily smokers (Johnson *et al*., 2002). Other genetics research has linked DSM-diagnosed nicotine dependence with the CYP2E1 genotype, which codes for a protein that metabolizes alcohol and tobacco smoke nitrosamines, and is implicated in creating metabolic cross-tolerance between alcohol and tobacco (Howard *et al.,* 2003).

Summary: There is evidence that diagnostic measures effectively index smoking heaviness, smoking-related health and mental health risks, and likelihood of future cessation. There is also strong evidence of heritability of DSM-diagnosed nicotine dependence. It is unclear whether paper-pencil versions of such measures (the TDS) are comparable to the interview versions of such measures. Moreover, there is evidence that the FTND may predict cessation and health outcomes as well as the diagnostic measures (e.g. John *et al*., 2004a). In terms of the prediction of likelihood of future cessation, it is unclear that diagnostic measures possess any incremental validity relative to briefer measures, such as the HSI. The diagnostic scales have relatively high reading levels, which may hinder their use with certain populations (even if administered orally).

Cigarette Dependence Scale

The Cigarette Dependence Scale (CDS) is another unidimensional tobacco dependence measure (Etter *et al*., 2003b). This assay was developed using smokers' reports of signs that they believed indicated addiction to cigarettes. Both a 5- and 12-item version of the CDS were developed (see Appendix 5). The items overlap somewhat with the Fagerström tests (e.g. they both assess number of cigarettes smoked per day and time to first cigarette in the morning). The Flesch-Kincaid Reading Grade Levels were 4.9 for the CDS-12 and 6.8 for the CDS-5.

Reliability and structure: To date, only two published studies have reported data on the two versions of the CDS, using data collected via the mail or Internet (Etter *et al.,* 2003b; Etter, 2005). The CDS-12 had strong internal consistency, the CDS-5 was within the acceptable range, and both scales were slightly skewed toward higher values. Test-retest correlations were ≥ 0.60 for all items and ≥ 0.83 for the full scales. Factor analysis suggested a unidimensional structure for the CDS-12.

Validation: The CDS scales were significantly correlated with number of cigarettes smoked per day (whether a smoker was a daily or occasional smoker), strength of urges during the last quit attempt, and cotinine level (Etter *et al*., 2003b). Curiously, the CDS-5 was more strongly correlated with cotinine levels than was the CDS-12. This was probably due to the fact that the question about smoking heaviness (Question #2) determined a greater portion of total scale variance in the 5-item version. In one study, none of the three dependence measures (i.e. the FTND, CDS-5, or CDS-12) was a significant predictor of relapse likelihood (Etter *et al*., 2003b); however, only a third of potential respondents participated in the follow-up study, which might have produced considerable response bias. In a second study, the CDS-12 weakly predicted smoking abstinence at 1-month post-quit, but in a counterintuitive direction (e.g. higher CDS-12 scores predicted abstinence) (Etter, 2005).

Heritability: To date, no data regarding heritability or genetics have been published using the CDS.

Summary: While the CDS scales do index smoking heaviness well, there is little evidence that they

predict likelihood of cessation effectively, or that they index other health outcomes of public health importance. Further, there is little evidence that they possess incremental validity relative to other measures, such as the diagnostic measures or the FTND. Overall, this measure is promising in that it can be used with paper-pencil administration and it has good reliability, but a meaningful evaluation must await additional validity research.

Multidimensional Measures of Tobacco Dependence

Nicotine Dependence Syndrome Scale

The Nicotine Dependence Syndrome Scale (NDSS) (Shiffman *et al.*, 2004) is a 19-item multidimensional scale based on Edwards and Gross' 1976 theory of the alcohol dependence syndrome. The NDSS was intended to complement, not replace, traditional dependence measures, such as the DSM-based assessments, and therefore there is little content overlap between the NDSS and the unidimensional measures. The NDSS assesses five dimensions of nicotine dependence: "Drive" reflects craving, withdrawal, and smoking compulsions; "Priority" reflects preference for smoking over other reinforcers; "Tolerance" reflects reduced sensitivity to the effects of smoking; "Continuity" reflects the regularity of smoking rate; and "Stereotypy" reflects the invariance of smoking (Appendix 6). The Flesch-Kincaid Reading Grade Level is 7.7. This reading level is somewhat elevated relative to other self-administered scales, which may reflect the fact that some items contain unusual words and require integration of more than one sentence or statement. For instance, the item, "My smoking pattern is very irregular throughout the day. It is not unusual for me to smoke many cigarettes in an hour, then not have another one until hours later," involves three negatives over its two sentences. In addition, some questions are double-barrelled, such as "It's hard to estimate how many cigarettes I smoke per day because the number often changes." If a person answers no, it is unclear whether the answer refers to difficulty of estimation per se, or because the number of cigarettes smoked per day does not change. Some items may be significantly influenced by cultural factors, such as eating in restaurants that are smoke-free or experiences during air travel. These features may make the NDSS somewhat less appropriate than some other measures for individuals of modest reading abilities or educational status. The NDSS has been translated into Finnish (Broms *et al.,* 2007).

Reliability and structure: To date, four studies of adult smokers have generated data on the NDSS; one study has reported on the NDSS in adolescents aged 12-18 (Clark *et al.*, 2005).

Psychometric data discussed here are based on the revised 30-item scale. The internal consistency for the NDSS total scale, the NDSS-T, is good (Shiffman *et al.*, 2004); however, data show that the internal consistencies of individual subscales are problematic (Piper *et al.*, 2006). Principal components analysis revealed a 5-factor structure for the NDSS (Shiffman *et al.*, 2004) as predicted by the underlying theory. Significant differences in the scores on the subscales between White and African-American smokers suggest the scale may operate differently in subpopulations, although there were no ethnic differences in the total NDSS score (Shiffman *et al.*, 2004). A more recent study, using the 19-item questionnaire with the Finnish Twin Cohort Study population, found that a 3-factor structure (priority/drive, continuity/stereotypy, and tolerance) best fit the data, with the internal consistencies of the three factors ranging from 0.83 to 0.92 (Broms *et al.*, 2007).

Validation: Much of the initial validation work was done with the 30- and 23-item NDSS, prior to its being refined to the 19-item version. These results indicated that the NDSS-T predicted time to lapse and time to relapse, but no individual subscale predicted lapse or relapse (Shiffman *et al.*, 2004). However, new data suggest that the NDSS subscales are significantly, though modestly, related to cigarettes smoked per day ($r = 0.12$-0.26) and that the Tolerance and Continuity subscales are modestly related to CO

level (r = 0.12 and 0.13, respectively) (Piper, *et al*., 2008). In samples of treatment-seeking smokers, the NDSS Priority and the Stereotypy subscales were found to predict cessation outcomes for up to 6-months post-quit (TTURC Tobacco Dependence Phenotype Workgroup, 2007; Piper, *et al*., 2008). The NDSS Drive, Tolerance, and the total score were found to predict heaviness of smoking and cessation outcome in a population-based sample of college students (Sledjeski *et al*., 2007). In Finnish smokers, the NDSS was significantly correlated with both FTND and DSM-IV, as assessed by the CIDI measures of dependence (Broms *et al*., 2007). The NDSS subscales accounted for 51% of the variance in self-reported difficulty abstaining among "chippers" (light/non-daily smokers) (Shiffman & Sayette, 2005), with the Drive subscale having the strongest relation (β = 0.61), relative to the other scales (β = 0.13-0.28).

Heritability: In the Finnish cohort, NDSS was found to have a significant heritability estimate of 0.30, relative to a heritability estimate of 0.40 for the FTND (Broms *et al*., 2007).

Summary: Like the CDS, the NDSS is a relatively new scale and it is not yet possible to draw firm conclusions about its validity relative to other dependence instruments. In its favour is the fact that some of its subscales have been shown to predict smoking heaviness measures, other dependence measures, and smoking cessation likelihood (Broms *et al.,* 2007; Piper *et al.,* 2008). The majority of this research has been done on clinical populations and it is not known how well these results would generalize to population-based samples. There is evidence that the various subscales of the measure are differentially related to various dependence criteria (Shiffman & Sayette, 2005; Broms *et al*., 2007; TTURC Tobacco Dependence Phenotype Workgroup, 2007). This suggests that some of the subscales possess discriminative validity with respect to different dimensions or aspects of dependence. However, there is evidence that the NDSS is not able to predict the major dependence criteria of smoking heaviness or cessation likelihood better than shorter measures (TTURC Tobacco Dependence Phenotype Workgroup, 2007). In addition, the marginal reliabilities of some of the subscales, and the reading level and complexity of some of the items, may discourage use in large population-based samples.

Wisconsin Inventory of Smoking Dependence Motives

The Wisconsin Inventory of Smoking Dependence Motives (WISDM) (Piper *et al*., 2004) is a 68-item measure developed to assess the discrete motivational basis of dependence. This measure has 13 theoretically-based subscales designed to tap different smoking dependence motives: Affiliative Attachment, Automaticity, Behavioral Choice/ Melioration, Cognitive Enhancement, Craving, Cue Exposure/Associative Processes, Loss of Control, Negative Reinforcement, Positive Reinforcement, Social and Environmental Goads, Taste and Sensory Properties, Tolerance, and Weight Control (see Appendix 7 for the items and scoring). The Flesch-Kincaid Reading Grade Level is 4.6; however, balanced against this easy reading level is the fact that the total scale is quite long. Therefore, investigators might wish to use individual, theoretically targeted subscales in epidemiologic research (subscales range from 4-7 items) (Lerman *et al*., 2006). Finally, relatively subtle psychological concepts are addressed in this measure, such as thinking of cigarettes as a friend or experiencing a loss of control, and this may affect the validity of such items in some cultures. There are English and Spanish versions of the WISDM (D.W. Wetter, personal communication, December 12, 2006).

While all subscales assess dependence, it should be noted that some of the subscales (i.e. Cue Exposure/Associative Processes, Social/Environmental Goads, and Taste/Sensory Properties) represent early-onset motives, which are present for all smokers even at modest levels of smoking experience, while other subscales represent late-onset motives (i.e. Affiliative Attachment, Automaticity, Behavioral Choice/ Melio-

ration, Cognitive Enhancement, Craving, and Tolerance), which are present only in individuals who smoke at a moderate daily rate or have at least moderate smoking experience (Piper *et al*., 2004).

Reliability and structure: To date, only one study has published data on the WISDM (Piper *et al*., 2004). Across two different samples all 13 subscales had strong internal consistencies that were evident across gender and across Whites and African-Americans. A new study found that the internal consistency of the subscales ranged from 0.74-0.94 with the total scale having a Chronbach's alpha of 0.96 (Piper *et al*., 2008). Factor analytic strategies indicated that the WISDM-68 is multidimensional, although some scales hit on related or overlapping dimensions of dependence. Thus, it is safe to say that some of the subscales are tapping the same underlying dimensions.

Validation: The total WISDM was correlated with smoking heaviness (cigarettes per day r = 0.63; CO r = 0.55) (Piper *et al*., 2004). Data also indicated that WISDM Total predicted outcome at both 1-week and 6-months post-quit (TTURC Tobacco Dependence Phenotype Workgroup, 2007). Thus, there is evidence that the whole scale is meaningfully related to the major dependence criteria. However, as with the NDSS, it appears that some shorter measures, such as the HSI, predict smoking heaviness and cessation likelihood as well or better than the longer WISDM (TTURC Tobacco Dependence Phenotype Work-group, 2007).

The various WISDM subscales show different patterns of relations with the dependence criteria. For instance, the Tolerance subscale was the best predictor of CO level, but the Craving, Cue Exposure/ Associative Processes, and Tolerance subscales were the best predictors of DSM-IV dependence when entered together into a multiple regression equation (Piper *et al*., 2004). One study found that although the total score was not a significant predictor of relapse after controlling for treatment, the combination of Automaticity, Behavioral Choice/-Melioration, Cognitive Enhancement, and Negative Reinforcement subscales all predicted relapse by the end of treatment in a multivariate model (Piper *et al*., 2004). Data from two different smoking cessation trials found that WISDM Automaticity and Tolerance were predictive of outcome at 6-months post-quit (TTURC Tobacco Dependence Phenotype Workgroup, 2007).

Heritability: There is evidence that the Taste/Sensory Properties subscale was significantly related to a genetic variant that determines sensitivity to bitter tastes (the phenylthiocarbamide (PTC) haplotype) (Cannon *et al*., 2005). Data have also revealed a significant relation between the WISDM Tolerance subscale with the ratio of 3-hydroxycotinine to cotinine (Piper *et al*., 2008). These data suggest that some WISDM subscales may code for biological diversity so as to permit genetic mapping.

Summary: Like the CDS and the NDSS, the WISDM is a relatively new scale and it is too soon to draw firm conclusions about its validity relative to other dependence instruments. However, data reveal that some of its subscales predict smoking heaviness measures and smoking cessation likelihood (Piper *et al*., 2008). There is also evidence that the various subscales of the measure are differentially related to various dependence criteria (TTURC Tobacco Dependence Phenotype Workgroup, 2007; Piper *et al*., 2008), suggesting that this measure is able to capture different dimensions or aspects of dependence. However, there is evidence that the WISDM is not able to predict the major dependence criteria of smoking heaviness or cessation likelihood better than shorter measures (TTURC Tobacco Dependence Phenotype Workgroup, 2007). Some WISDM subscales have been related to various dependence-linked genetic components. It is important to note that the WISDM research has been done on clinical populations and it is not known how well these results would generalize to population-based samples.

Summary:

Assessment of cigarette-induced nicotine dependence is an important goal for three reasons. First, the human and economic

costs of cigarette-induced, nicotine dependence is significant. Second, only a portion of cigarette smokers are "dependent" (as defined by traditional instruments), and those who are dependent are indeed distinguishable from other smokers on the basis of factors, such as likelihood of future cessation and amount smoked daily. Finally, cigarette-induced nicotine dependence may serve to moderate individuals' responses to different tobacco control programmes and policies, as well as the proximal and distal effects of these interventions.

It is important to note that there is considerable evidence that the various measures of nicotine dependence are not highly related to one another, and can have very different relations with validity measures (Hughes *et al.,* 2004a; Piper *et al.*, 2006). Thus it is critical that investigators select measure(s) that are psychometrically sound, appropriate for the intended population, and target the constructs in which the researchers are interested. If the goal is to assess a central core of nicotine dependence as a predictor of cigarette use cessation likelihood, or as an index of associated health risks, then the FTND or HSI appear best suited for this purpose (Tables 3.22 and 3.23). These instruments are brief and have relatively impressive predictive validities, and their reading level should make them appropriate for a broad range of populations. However, it is important to note that none of the dependence measures accounts for a large proportion of variance in outcomes in cessation likelihood. This is no doubt due to the fact that cessation likelihood is affected by countless situational/environmental factors, and other person factors. In addition, if one uses a brief measure, such as the HSI, it is important to recognize that it does not tap all dependence factors. It also does not appear to predict certain core features of dependence well, such as withdrawal, and it may be inappropriate in populations that do not smoke daily or have significant restrictions on smoking (e.g. restrictions that constrain smoking in certain contexts or times of day).

There may be situations when there is a need to assess particular, relatively discrete, facets of nicotine dependence. For example, identifying specific tobacco dependence mechanisms may facilitate: identification of a more proximal phenotype (Cannon *et al.*, 2005), identification of specific dependence dimensions with which one could create treatment algorithms, monitoring of the development of tobacco dependence, or identification of a specific group of dependent tobacco users for whom a policy is particularly effective or ineffective. If this is the goal of the research, then a multifactorial measure (i.e. the NDSS and the WISDM-68, and their subscales) would be optimal, despite the fact that there is little evidence for incremental validity in predicting important public health outcomes. However, the relative lack of validity information on these scales may mean that researchers should use these instruments only in the context of exploratory research. They might be most appropriate for research addressing etiology and cultural or population-based differences in smoking determinants.

Measures of smokeless tobacco-induced nicotine dependence

Like cigarettes, smokeless tobacco (ST) products contain nicotine, although the levels vary considerably across products (Hatsukami *et al.*, 1992; IARC, 2007b). Data on patterns of use of ST support the conclusion that many users are nicotine dependent (Henningfield *et al.* 1997; IARC, 2007b). Many ST users experience withdrawal symptoms upon abstinence (Hatsukami *et al.*, 1992; 1999). Studies have used a biomarker of nicotine uptake, cotinine, to show that daily users of ST exhibit levels of nicotine absorption that are equivalent to daily cigarette smokers (Gritz *et al.*, 1981).

Dependence on smokeless tobacco has often been assessed with questionnaires derived from FTND, with the addition of specific items, in particular, swallowing the tobacco juice (Boyle *et al.*, 1995; Ebbert *et al.,* 2006). In three different samples, use of ST within 30 minutes of waking and swallowing the tobacco juice were

Construct	Tobacco Dependence
Measure 1	Fagerström Test of Nicotine Dependence (FTND) – 6 items
Source	Heatherton *et al.*, 1991
Variation	It is possible to change the wording of the items to be culturally appropriate or to reflect non-cigarette tobacco use. However, these changes may affect the reliability and validity of the data obtained.
Validity	• Predicts both behavioural (e.g. lifetime amount smoked) and biochemical (e.g. CO, cotinine) indices of smoking in multiple countries • Predicts cessation • Evidence of linkage to specific dependence-linked genetic variants
Comments	This measure is recommended as an assessment of dependence's ability to predict cessation and heavy use • Brief and well-known • Strong predictive validity of heavy use and cessation • Internal consistency is modest, which may reflect a 2-factor structure • Some items may be influenced by smoking restrictions in the environment • Has been translated into a number of different languages
Measure 2	Heaviness of Smoking Index (HSI) – 2 items from the FTND: number of cigarettes smoked per day and time to first cigarette in the morning
Source	Kozlowski *et al.*, 1994
Variation	It is possible to change the wording of the items to be culturally appropriate or to reflect non-cigarette tobacco use. However, these changes may affect the reliability and validity of the data obtained.
Validity	• Predicts both behavioural (e.g. lifetime amount smoked) and biochemical (e.g. CO, cotinine) indices of smoking in multiple countries • Predicts cessation – the HSI appears to be the strongest predictor of cessation, accounts for much of the predictive validity of the FNTD • Highly heritable (71%) and linked to specific dependence-linked genetic variants
Comments	This measure is recommended as the most efficient measure to assess dependence's ability to predict cessation. • Brief • Using this measure may only involve the addition of item (time to first cigarette) if number of cigarettes per day is already being collected • Strong predictive validity of heavy use and cessation • Items may be influenced by smoking restrictions in the environment • Has been translated into a number of different languages

Table 3.22 Measures of Cigarette-Induced Nicotine Dependence

Construct	Tobacco Dependence
Measure	Fagerström Test of Nicotine Dependence (FTND) – 6 items
Sources	Boyle *et al.*, 1995; Ebbert *et al.*, 2006
Variation	It is possible to change the wording of the items to be culturally appropriate or to reflect non-cigarette tobacco use. However, these changes may affect the reliability and validity of the data obtained.
Validity	• Predicts both behavioural (e.g. lifetime amount smoked) and biochemical (e.g. CO, cotinine) indices of smoking in multiple countries • Predicts cessation • Evidence of linkage to specific dependence-linked genetic variants
Comments	This measure is recommended as an assessment of dependence's ability to predict cessation and heavy use • Brief and well-known • Strong predictive validity of heavy use and cessation • Internal consistency is modest, which may reflect a 2-factor structure • Some items may be influenced by smoking restrictions in the environment • Has been translated into a number of different languages

Table 3.23 A Measure of Smokeless Tobacco-Induced Nicotine Dependence

the variables most consistently associated with cotinine level (Boyle *et al.*, 1995) (see Appendix 8 for the items and scoring).

Summary:

Like cigarettes, smokeless tobacco can result in nicotine dependence. While less research has been done to validate self-report measures of ST-induced nicotine dependence, questionnaires derived from FTND appear to provide a means for identifying ST users who are nicotine dependent.

Summary and recommendations

Nicotine dependence is an important construct to assess as a moderator for the effects of tobacco control programmes and policies. In this section the evidence was reviewed on the validity of various proposed measures of cigarette and smokeless tobacco nicotine dependence. For cigarette smoking, the 2-item Heaviness of Smoking Index is recommended for use in population level studies. If only a single item measure is possible, the use of "time to first cigarette in the morning" is recommended. For smokeless tobacco, the FTND-ST appears to be a useful measure of nicotine dependence.

4.1 Data sources for monitoring tobacco control policies

Introduction

Do we know why the prevalence of smoking in Sri Lanka decreased from 54% in 1988 to less than 40% in 2003? What is this decrease related to? Does tobacco control have a part in this? If so, what specific policy interventions were most useful in decreasing the prevalence in Sri Lanka? How does that compare to other countries? To respond to these and similar questions on the relationship between the implementation of specific tobacco control policies and tobacco use prevalence in any country, researchers and policy-makers need a solid understanding of the current state of policies and their specific impact at the country level (http://www.who.int/ncd_ surveillance/infobase/web/InfoBaseCommon/).

This section describes the currently available sources of information on tobacco control policy interventions, with special attention to the new WHO Global Tobacco Control Report system, and assesses their credibility, completeness, and usefulness. It also discusses important methodological issues and gauges future prospects for such systems. Although tobacco control policy interventions can be initiated by private sectors of the civil society, this section is concerned with core governmental policy interventions, since in most countries, only governments have a population-wide reach and the capacity and authority to consistently enforce stringent measures. Such interventions typically include any governmental form of regulation, funding decision, institutional statement, organisational development, or administrative action to apply (or not apply) tobacco control policies. Further down, this section discusses evaluation criteria for tobacco control policy interventions monitoring systems, and reviews currently available data sources based on these criteria. The last part of the section builds on the first two and discusses renewed efforts to build comprehensive tobacco control monitoring systems in the new international tobacco control context.

Criteria for assessing tobacco control policy intervention monitoring systems

An ideal global tobacco control monitoring system would track interventions to decrease tobacco use in all relevant policy domains, and would make the data comparable across all jurisdictions, based on an explicit and transparent protocol. Such a monitoring system would have the following characteristics: (1) include all relevant tobacco control policies and regularly be updated to include new innovative policies; (2) characterise the interventions against current best practice standards; (3) include the degree of enforcement of policy interventions; (4) rely on credible sources; (5) cover all countries, as well as all relevant sub-national jurisdictions; (6) be updated as changes occur, or at least at regular and short intervals, while keeping historical information; and (7) span a long enough period to link changes in tobacco control policies to changes in the prevalence of tobacco use and other impact indicators. Therefore, tobacco control monitoring systems are assessed in this paper in relation to the following variables:

- Policy scope
- Characterization of interventions against best practice standard
- Characterization of degree of enforcement
- Source of the primary data
- Geographical/jurisdictional coverage and comparability of data across jurisdictions
- Timeliness and frequency of data collection
- Characterization of evolution of policies over time.

Scope of policies covered:

Tobacco control interventions are wide in scope and vary in time and space. However, despite the sheer diversity of possible policy interventions, they can be regrouped in a few convenient categories that generally fall under "demand reduction" measures and "supply reduction" measures (although some policies do not easily yield to this rather strict dichotomy) (Table 4.1). In assessing the scope of tobacco control data systems, one must bear in mind that not all tobacco control policies are equal. Supply reduction policies are generally considered not to be very effective at reducing tobacco use, except perhaps for anti-smuggling measures under certain conditions (Rowena *et al.*, 2000). Given the limited resources devoted to data gathering, efforts should first be dedicated to demand reducing policies. Even among such policies, however, there are wide differences in effectiveness. Increasing prices through high taxes, as well as smoke-free environments, are generally seen as the most effective tobacco control policies (Ranson *et al.*, 2000), and therefore are considered essential in any information system.

If resources allow, clearly ineffective policies could be monitored. This could provide a scan of the policy environment and assess the imbalances produced by focusing on ineffective measures. For example, in the context of constant aggressions from the tobacco industry to avoid effective tobacco control, monitoring measures that are inefficient, but at the same time (and for the same reason?) the darling of the industry, might indicate how misguided the policy priorities of a given jurisdiction are. Examples are the effectiveness of school-based education programmes and prohibition sales of tobacco products to minors (Ling *et al.,* 2002; Cummings *et al.*, 2003; Glantz *et al.*, 2005; Wiehe *et al.*, 2005). Tobacco control monitoring systems should be assessed on the strategic choice of the policy domains and interventions they cover. Although this choice is generally implicit in all datasets, the data collector should clearly describe the basis for that choice, whether in terms of efficacy, or any other criteria.

Characterization of tobacco control policies based on best practice standards:

Once the scope is established, the data system must be assessed in relation to its capacity to characterise each policy according to an explicit standard or recognised best practices. For example, it is generally acknowledged that bans on advertising, promotion, and sponsorship should be comprehensive. Therefore, systems monitoring marketing restrictions should be

Demand Reduction Policy Domains	Supply Reduction Policy Domains
Price and tax of tobacco products	Liability and litigation
Protection from exposure to secondhand tobacco smoke	Access to tobacco by youth
Tobacco advertising, promotion and sponsorship	Banning sales of tobacco products
Packaging and labelling of tobacco products	Crop substitution
Treatment of tobacco dependence	Contents of tobacco products
Education, communication and public Awareness	Illicit trade in tobacco products

Table 4.1 Tobacco Control Interventions

assessed on their ability to provide information that would allow gauging the policies of any given jurisdiction against this standard.

According to this standard, the monitoring system must then select all relevant variables describing the components of this policy and collect the data accordingly (Joossens *et al.,* 2006). Following with the previous example, to ascertain the existence of a comprehensive ban on advertising, promotion, and sponsorship, the monitoring system should provide information separately on each form of commercial communication, recommendation or action, and any form of contribution to any event, activity, or individual with the aim, effect, or likely effect of promoting a tobacco product or tobacco use either directly or indirectly. Such a policy would include data on the existence of direct advertising bans of tobacco products or brands in every existing media, including national and international TV from any source (cable, satellite, internet, etc.), national and international radio, local and international magazines and newspapers, billboards, points-of-sale, the internet, and cinemas. Moreover, the monitoring system should collect data on the ban of each specific form of promotion of tobacco products, brand names, or company names, including direct mail giveaways, promotional discounts, non-tobacco products identified with cigarette brand names, brand name of non-tobacco products used for tobacco product, product placement in TV and/or films, and sponsored events. In addition, the existence of each element of the policy should be assessed with a Yes/No question that leaves little room for interpretation and explicitly meets the best practice standards. The monitoring system should clearly describe the criteria used in answering Yes/No questions, and these criteria should be termed in the same language as the laws. For instance, these apply notably in deciding whether a smoking ban is complete, whether health warnings are effective, whether advertisements are banned from the media, etc.

To have a clear characterization of any given policy intervention is not always easy. Even with all necessary legal information, the data collector is left to match their own definition of the desired policy with the jargon of the law. One desired policy might be a complete smoking ban. However, even good laws typically do not provide for “complete” bans and could include some exemptions. The Irish law is a case in point; it does not provide for a complete ban strictly speaking. However, judging when exemptions are minor or not might be a challenge, and setting a clear and detailed standard of excellence is important in assessing and collecting data for monitoring tobacco control policies. More complicated is the assessment of the Italian law. It does contain a smoking ban, but exceptions are allowed in the form of smoking rooms, usually not considered a best practice. If the applied standard considers that bans with smoking rooms are not complete, the Italian ban would not complete. However, the requirements for smoking rooms are so stringent, that Italian law *de facto* can be considered providing for a complete smoking ban, as smoking rooms are rarely available.

Characterizing of any given policy intervention becomes even more difficult in the absence of clear information on regulations. Some countries have legal systems where regulation is very general, leaving it to administrative actions to determine how regulations are to be applied. Some regulations have loopholes; some countries have contradictory decrees issued by many types of authorities, with uncertain rules determining which decree has precedence. In other countries, one must consider jurisprudence and court orders suspending or modifying regulations.

In summary, any tobacco control monitoring system, because it attempts to verify the existence of an implicitly defined “good" policy intervention, must synthesize complex information to answer simple questions. At one time or another, collecting the information may call for some judgment by the data collector. A good tobacco control monitoring system should minimize the impact of these judgment calls and make them as explicit as possible.

Enforcement

Any characterization of a policy intervention is not complete without assessing the actual enforcement of the measure. It is not enough to know that a policy intervention legally exists without knowing if it is applied. The system monitoring tobacco control policies can use two types of measures to assess the enforcement of a policy intervention: *de facto* implementation of the intervention in conformity with the policy, and enforcement efforts by the government. The first type of measure is best since it addresses exactly what needs to be gauged, while the second method is an indirect indicator that looks at the process leading to enforcement.

De facto implementation requires specific quantitative metrics based on direct observation of people or events, outside the purview of a monitoring system. Such measures are often unrealistic for many countries with low resources; measuring enforcement of smoking bans, for example, may require population surveys, sometimes including biological measures of exposure to secondhand smoke. Other metrics might include data provided by the industry, because of clear legal obligations (e.g. detailed sales or advertising data), that can help understand the impact of policy. Although preferable to other approaches, direct observation is not exempt from problems. Even surveys are difficult to interpret. In Brazil, for example, 70% of respondents to a survey reported that they had seen billboards with tobacco advertisement in the month before the survey, despite a successful complete ban enforced 5 years earlier. It is thus possible that survey respondents did not understand the question or that they might actually be reporting types of advertisement that are not covered by the law (Global Youth Tobacco Survey fact sheets; http://www.cdc.gov/tobacco/global/GYTS/factsheets/paho/factsheets.htm).

A more feasible alternative is to rely on the opinion of key informants or experts, providing some sort of qualitative direct observation. The panel of key informants or experts is especially sensitive to judgment calls and must be assessed very carefully. In this respect, developing a stringent, multi-layered protocol is probably a sound base, but there is not yet a consensus on what would be a method that is inclusive enough at the national level, yet comparable enough at the international level. Indeed, qualitative assessment of enforcement is not easy, especially at the international level, where national experts might have a widely different appreciation of enforcement.

Methods based on quantitative measures can be used to gauge efforts (usually by the government) leading to enforcement. These can be measured by enforcement budget, number of full-time equivalent inspectors, number of inspections, number of fines distributed, etc. There are two obvious problems with such measures: first is that the existence of enforcement efforts does not indicate enforcement of the law necessarily; and second, the absence of enforcement efforts is not an indication of lack of enforcement in countries where tobacco control measures are widely respected without severe enforcement. Countries where interventions are self-enforcing from the beginning, or where significant efforts might not be needed after many years of successful enforcement, will fare quite badly next to a country with a severe enforcement problem despite significant government efforts. In addition, such statistics are not always available. In fact, in some countries, it is not clear who should enforce the law, and gathering statistics then becomes difficult. In the case of smoke-free environments, for example, sometimes police are in charge of enforcement and often do not use fines to enforce the law, given the low social acceptability of a fine for smoking in a restaurant; casual reprimand is used instead and no trace is left in any official record.

Given these difficulties and inherent limitations of the second approach of measuring enforcement efforts, it is probably better to mainly rely on the first approach, but to also use some basic measures of government efforts that are in line with recommendations on enforcement. Monitoring systems could, in this case, gather data on the existence of a clearly identified body in charge of enforcing the law, and if

possible, the budget or staff of that specific agency or unit, if it exists.

Whichever approach is used, a monitoring system should be assessed on its explanation of the measure of enforcement used. The choice of approach and method must thus be explicit. If it uses a survey, a panel of experts, or any other investigation method to determine the actual impact of a policy, this method must be described in detail so that the reader can clearly understand it strengths and limitations.

Source of the primary data

The scope and characterization of policy interventions described above are key to assess the relevance of the contents of an information system. However, the crucial element to evaluate the quality of the information it provides is the assessment of the primary source of data. Written laws and regulations are the usual source of primary data for policy interventions. Monitoring systems should make all legal documents available for users to consult when in doubt (online if possible), so that the reader can see what relevant information was collected.

However, assessing the existence of some policy interventions cannot be done by looking at the written regulations. This is typically the case of treatment and education efforts. The presence of an easily reachable quitline, for example, requires a measure of actual existence and use. Observing and characterizing these policy interventions must often rely on surveys, *ad hoc* metrics, qualitative measures, and expert judgment. Moreover, it is very difficult to use a method that is suitable for all national contexts; hence, the importance of describing and justifying methods used.

Geographical/jurisdictional coverage

An ideal monitoring system should provide data on policy interventions in all countries of the world, and in all relevant subnational jurisdictions within each country. Worldwide geographical coverage comes at a cost; a balance must be struck between coverage and thoroughness. Not only can resources prove to be a constraint, but the wider the geographical coverage, the more difficult it becomes to make the data comparable, and the less uniform relevant policy scope tends to be. The goal of the monitoring system must thus be carefully considered before deciding what the best geographical coverage is.

In general, global coverage should be the main goal, with very clear questions and definitions and thought to specific regional issues. Given the broad diversity in national contexts, this type of exercise should also be decentralized; hence, the necessity for a wide, yet highly coordinated, network in order to make the data comparable. Such focus, however, should not preclude the existence of regional variations over and above a common core set of questions, in order to increase flexibility of the exercise and country level relevance and buy-in.

The coverage of specific subnational jurisdictions follows the same principle. In the countries where this is relevant, inclusion is an absolute priority. In Canada, for example, very stringent smoke-free laws are enforced at the provincial level, and excluding provinces would result in faulty answers. Yet, there are only a few cases where inclusion of subnational jurisdictions is essential, and once more, local knowledge on the existence and relevance of these policies is critical. Should municipal by-laws be included for example? What if a city comprises a significant minority or even a majority of the population and has such by-laws? Given the complexity of some political systems and jurisdictions, this will typically require local consultation. These questions can only be resolved on a case-by-case basis, hence the necessity of the monitoring system to outline clear guidelines for inclusion/exclusion. Among the guidelines is the stability of these institutions and laws, number of people affected by the laws, their share in the national population, strong within-country variations, etc.

Timeliness and frequency of data collection

Given the pace of change in the field of tobacco control, an ideal monitoring system should be live, that is, updated as changes occur. Live systems demand the

existence of a stable tobacco control country level network and a central coordination mechanism. Short of that standard, and in the absence of a stable network, the frequency of updates should mainly depend on budgetary issues, with a careful balance to be struck between the frequency of updates and budgetary sustainability. In all cases, the data should not be more than one or two years old, or the time it takes for these policies to significantly affect prevalence.

Change of policies over time

Old data should also be kept and made available, so that researchers can track the evolution of policy in an attempt to link it to prevalence. Old laws, date of changes in the law, date of changes in the implementation of the law, etc., are all very important for monitoring systems whose aim is to track the evolution of policy, and not just current policy, if we are to assess these measures.

Description and assessment of current data collection systems

Only two global tobacco control monitoring systems are presently operational: the WHO Global Tobacco Control Report (GTCR) and the reporting instrument of the Conference of the Parties (COP) to the WHO FCTC. The GTCR is based on the previous work of the National Tobacco Information Online System (NATIONS) and on still existing WHO regional databases. Described below are the reporting instruments of the WHO FCTC, the precursors of the GTCR, and the GTCR itself.

The reporting instrument of the Conference of Parties to the WHO FCTC:

The WHO Framework Convention on Tobacco Control (WHO FCTC) is the first treaty negotiated under the auspices of the WHO. It was adopted unanimously at the 56th World Health Assembly, in May 2003. Its provisions obligate only parties that have ratified the treaty, which as of September 2008 were 160 WHO member states. An important provision of the WHO FCTC is that each Party is obligated to submit periodic reports on its implementation of the Convention, in accordance with Article 21. To this end, the first meeting of the COP in 2006 provisionally adopted a reporting system whose objective is to understand and learn from the various experiences of parties in implementing the WHO FCTC. Questions in the reporting instrument are clustered into three groups. Only Group 1 questions have been designed and applied by countries reporting to the second meeting of the COP in 2007 [the third meeting of the COP on November 2008 approved changes to Group1 questions].

Scope and characterization of interventions:

Given the need to report on the wide range of obligations contained in the WHO FCTC, the policy scope of the COP reporting instrument is very large, but does not directly prioritize policies in terms of effectiveness. This instrument contains "Group 1" questions, which are wide in scope and range from tobacco use prevalence to measures taken to curb illicit trade, as well as education, and public awareness programmes. Core Group 1 questions require information about tobacco use, licit supply of tobacco, duty-free sales volume, price and tax measures to reduce demand for tobacco, regulation of tobacco product disclosures, illicit trade in tobacco products measures, seizures of illicit tobacco, education, communication, training and public awareness activities, measures on sales to and by minors, liability measures, management of tobacco dependence and cessation services, measures to support alternatives to tobacco growing, research, surveillance and exchange of information, programmes and plans, national coordinating mechanisms, and technical and financial assistance provided and received.

The data is collected at the country level, and its purpose is not to provide a uniform framework for comparison, but rather a way of observing the progress of the implementation of the treaty obligations within each country. Therefore the possibility of comparing answers across countries is extremely limited, although the questions on legislative measures are in general quite detailed.

Enforcement:

There are no enforcement measures considered in the COP reporting instrument.

Data sources:

The information is self-reported by governments, which are required to provide the supporting legislative documents. However, there is no external validation planned. The absence of any formal standardization process, beyond the instructions of the reporting instrument, might mean that the user should go back to supporting documents in a systematic fashion. This is especially the case for the questions regarding legislation, where countries are asked if they have "adopted and implemented legislative, executive, administrative, or other measures" on specific policies whose level of implementation is sometimes quite vague (e.g. smoke-free environments are defined as "full," "partial," or "none", without any specific definitions of these terms).

Geographical coverage:

The geographical coverage of the reporting instrument is limited to the signatory parties, although the number of parties increases regularly and might finally include all WHO member states. The issue of subnational legislation is also absent from the questionnaire.

Timeliness, frequency of data collection, and trend:

Group 1 questions must be answered within two years of entry into force of the Convention for that Party, and then every three years after that. Group 2 and 3 questions must be reported within five and eight years of entry into force of the Convention for that country, respectively. [Group 2 questions were approved in November 2008. However, Group 3 questions have not been designed yet]. By the end of 2008, 140 parties will all have completed the Group 1 questions for the first time.

The main goal of the reporting instrument is to report on treaty implementation and not on tracking the evolution of tobacco control. In this respect, following the trend of legislative measures is not an objective of the COP reporting instrument. The periodic reports submitted by parties, however, may allow some trend analysis within each country.

In summary, the WHO FCTC reporting system in its current form is not designed to be a thorough, scientifically-oriented, annual monitoring programme. It has serious limitations on the immediate use of its data for monitoring policy interventions and comparing legislative measures across countries. Once the data are available publicly, however, independent researchers can undertake the type of work they choose to, but it will be based on their own interpretation of the data and their own assumption on the comparability of the information, since there is no a detailed protocol to make the data comparable.

The reporting instrument, however, might evolve towards a monitoring system. An independent assessment of the current system is scheduled for 2009; the COP will further consider the matter of reporting in 2010. Already decisions of the second COP, that gathered in Bangkok in the summer of 2007, point to the need for increased standardisation through an improved questionnaire, as well as through the long-term evolution of the questionnaire with Group 2 and Group 3 questions.

The Global Tobacco Control Report (GTCR) precursors: NATIONS and the WHO regional databases

Although NATIONS (http://apps.nccd.cdc.gov/nations/) is not updated anymore, it was the first global monitoring system for tobacco control and played a historical role for later efforts. NATIONS was a collaborative effort by the United States Centers for Disease Control and Prevention (CDC) and the WHO, and also involved the American Cancer Society (ACS), and the World Bank (WB). Its aim was to monitor tobacco use and control, based on data gathered from several sources that stretched from governmental and international agencies to commercial entities, scientific literature, etc. A lot of the data was originally collected by the ACS and the

WHO to prepare the monograph *Tobacco Control Country Profiles,* which was first published in 2000, followed by a second edition in 2003 (Shafey *et al*., editors). After the adoption of the WHO FCTC by WHO Member States in May 2003, the data and further responsibility for collection efforts was transferred to the WHO, and they undertook the creation of regional databases through their regional offices.

The data gathering process also underwent important changes. Data collection was decentralized to the regional level in order to increase proximity to the countries and obtain more accurate information on tobacco control measures and their implementation. The data being collected through the WHO regional offices became official, and had to be validated by national authorities before it could be published. The WHO Regional Office for Europe (EURO)(http://data.euro.who.int/tobacco/) has so far provided the most comprehensive data collection effort and has the most complete regional dataset of all regional offices. This database is used in turn to support the *European Tobacco Control Report*, a publication with detailed information on the state of tobacco control in the 52 countries of EURO (http://www.euro.who. int/InformationSources/Publications/Catalogue/20070226_1). What follows is a description of the EURO database.

Scope and characterization of interventions:

The scope of policies covered in the EURO database is ample (Table 4.2). As for NATIONS, the data covers more than tobacco control (e.g. prevalence, mortality, economics of tobacco); it additionally covers policies, such as taxation and cessation.

The criteria for guiding the choice of policies are not explicitly provided, and the dataset includes tobacco control measures of very diverse cost-efficiency without characterizing them. The protocol and definitions to make the data comparable is also absent from the publicly available information on the website. This might lead to some comparability issues. In the case of smoke-free environments, for example, the situation of a country is classified into one of three categories: smoking bans, restrictions, and voluntary agreements. The first problem is that "smoking bans" in the EURO database are not really complete and might allow for some exceptions. The second problem is that "voluntary agreements" are not described to ascertain if, independently regulated by law of the agreement, they prescribe a 100% to smoke-free environments or not.

The same issue applies to all other tobacco control measures, where a clearer and more explicit protocol would be needed. The description of each tobacco control measure, and their characterization in terms of "Yes" and "No," are much more detailed than in NATIONS, thus leaving less room for interpretation by the data collector. The format of some of the data could also be improved, such as the tax data that provides not the rates, but the share of the price of a pack that goes into different types of taxes; the underlying tax rates and the methodology to convert them in share of the prices would be useful. However, most legal documents that were relied on are available on the website (except for taxes), thus mitigating that problem.

Enforcement:

The enforcement is assessed by the opinion of the focal point[1] collected by completion of a questionnaire. A score of 1 to 5 is provided for the enforcement of smoke-free legislation, bans on direct and indirect advertising, product regulation, and sales to minors. However, the assessment is not published on the website.

[1] A National Focal point (NFP) is a national centre, designated by each State Party, which is accessible at all times for communications with WHO International Health Regulation Contact Points. While the exact structure and organisation of the NFP are left to the State, IHR (2005) define the role, functions and operational requirements for real time management of information and for efficient communications. It is foreseen that NFPs will be offices rather than individuals.

Tobacco Use	• Smoking prevalence in adults • Smoking prevalence in young people
Economics	• Cigarette consumption • Cost (in money and labour) of tobacco products • Tobacco tax revenues from excise duties • Duty stamps, earmarking of tobacco taxes • Licensing • Government ownership and financial incentives • Studies of smuggling, economic and social costs, and litigation • Annual price variations of tobacco products in real terms (%) • Structure of taxation of tobacco products
Laws and Regulations	• Direct advertising of tobacco products • Indirect advertising of tobacco products • Distribution of tobacco products through various outlets • Regulations for sale of tobacco products • Smoke-free areas • Smoke-free public transport • Health warnings • Measurement, regulation and disclosure of tobacco product ingredients and smoke constituents • Treatment of dependence: - Interventions to support smoking cessation - Quitlines - Availability of smoking cessation treatment - Training for health professionals • General policy: different sub-national laws or regulations • Public information and advocacy • Participation in WHO networks
Health Consequences and Costs	• Average number of years lost per death from smoking (years) • Deaths attributed to smoking in all ages • Deaths attributed to smoking in middle age (35-69 years) • Proportion of deaths attributed to smoking in all ages (%) • Proportion of deaths attributed to smoking in middle age (35-69 years) (%) • Standardised death rate from trachea, bronchus, or lung cancer (per 100 000)

EURO: WHO Regional Office for Europe

Table 4.2 Scope of Policies Covered by the EURO Tobacco Control Regional Database

Data sources:

This database relies on a questionnaire that was distributed to national level tobacco control focal points, who often work from within their national Ministry of Health, thus ensuring accuracy and country endorsement. The data source is thus highly credible, but this process is not described on the website, so the reader cannot assess the validity of the information. Main sources are legislative measures to control tobacco, although other policies are also monitored, such as prevalence and epidemiological impact of tobacco consumption, as well as tobacco economics.

Geographical coverage:

The EURO database covers all European countries. Although data from subnational jurisdictions is not available, its existence is assessed for eight categories of legislative measures.

Timeliness, frequency of data collection, and trend:

The data collection involves a lot of back and forth between countries and the regional office, in order to clarify and standardize answers, as well as ensure country buy-in. This, however, creates long delays between initiation and conclusion of the data collection effort. The last round of data collection, for example, was initiated in June 2005, but was not completed until the fall of 2006, which allowed for potential inaccuracies for countries that legislated during this period. The process of updating the data is not specified and there is no built-in regular update mechanism.

Situations in other regions:

Not all regional offices had the means to set up systems as complete as that of EURO (http://www.who.int/tobacco/global_data/regional_databases/en/index.html). In the Africa Regional Office (AFRO), the system does not exist and the outdated NATIONS represents the main source of data. In the Eastern Mediterranean Region (EMRO; http://www.emro.who.int/TFI/CountryProfile-Part6.htm) and the South East Asia Region (SEARO; http://www.searo.who.int/), the data was compiled in 2000-2002 and has been updated in 2008. The policy scope is much narrower than in EURO, reasons for selecting the indicators are not specified beyond being “relevant and readily available,” and geographical coverage could be improved. As for other regions, the protocol or criteria for interpreting the laws is not explicitly described, thus raising issues of comparability between countries, but mostly between regions (some EMRO legal texts are available online). In the Pan American Health Organization (PAHO; (http://www.paho.org/tobacco/PatiosHome.asp) and the WHO Western Pacific Region (WPRO; (http://www.wpro.who.int/), the situation is somewhat in between that of EURO and the regions with least policy database documentation, and the datasets cover mainly the information available in legal texts for a subset of countries. Criteria for assessing this information are much more detailed, with very specific questions leaving little room for interpretation.

Overall, the WHO regional databases represented until now the best existing global data source on tobacco control policies. However, they suffer from many issues, of which timeliness and lack of enforcement data are the most immediately obvious ones. Most important is that the tobacco control indicators are not the same between regions, and are not defined with the same criteria (besides the fact that these criteria are never fully described). This raises serious issues of overall comparability.

The Global Tobacco Control Report (GTCR) system

The Global Tobacco Control Report (GTCR), released in early 2008, is the central instrument of a worldwide tobacco control monitoring effort by WHO (http://www.who.int/tobacco/mpower/en/). The objective of the report is to monitor a core of essential tobacco control policy initiatives, and to report on their implementation on an annual basis. The GTCR aims to provide a highly structured and focused framework through which progress towards the implementation of defined, concrete tobacco control measures at the country level will be

compared in a standardised manner across countries. Essential indicators are measured through a short questionnaire that is completed by country level focal points.

Scope and characterisation of interventions:

The GTCR focuses on a few policies that were selected based on their efficiency and cost-efficiency. The questionnaire requires information on national prevalence of daily tobacco use; the share of tobacco taxes in the price of a pack; the existence of visible health warnings occupying at least 30% of the package of tobacco products; complete advertising, marketing, and promotion bans of tobacco products by type of media; complete smoking bans by sector; the availability of tobacco dependence treatment; and existence of national tobacco control policy objectives. Policies such as awareness campaigns or anti-smuggling initiatives are not considered. Answers to this annual questionnaire will be analysed in the GTCR, which will use gaps between optimal and existing policies revealed in these data and analyses to develop a strong advocacy message. Table 4.3 provides the scope of policies covered by the GTCR.

Enforcement:

The GTCR uses the following protocol to assess the enforcement of smoke-free environments, as well as direct and indirect advertising bans for each country (Table 4.4). The assessment of enforcement is integrated globally through an enforcement score, where a highly enforced policy is worth two points, a moderately enforced policy one point, and a minimally enforced policy no points, hence a maximum score of 10 points given the five experts. This system, although very simple, works quite well with the majority of countries with legislation providing the assessment and enforcement scores conforming to expectations. Moreover, the scores are credible at the global level, with a wide dispersion of values, as well as within countries, with very few polarized expert assessments and yet very few consensual situations. The score, however, suffers from

Tobacco use	• Internationally comparable smoking prevalence in adults
Economics	• Structure of taxation of tobacco products • Earmarking of tobacco taxes • Tobacco tax revenues from excise duties • Price of main cigarette brands
Laws and Regulations	• Direct advertising of tobacco products • Indirect advertising of tobacco products • Smoke-free areas • Health warnings • Treatment of dependence: - Interventions to support smoking cessation - Quitlines - Availability of smoking cessation treatment • General policy: different sub-national laws or regulations

GTCR: Global Tobacco Control Report

Table 4.3 Scope of Policies Covered by the GTCR

1	Choose five key (non-paid) experts of different institutions and professions. Preferably select individuals with the following background: (1) one health professional with a strong background in tobacco control, (2) one academic who specializes in tobacco control, (3) the head of a prominent non-governmental organisation in tobacco control, (4) the government official responsible for tobacco control activities, (5) the WHO focal point for tobacco control (who usually is also filling out the questionnaire).
2	Consult the experts separately. In many countries, tobacco control networks are very small and the same individuals might wear many hats. For example, the chief tobacco control officers in the government are often dedicated to the point of also being the head of leading tobacco control non-governmental organisations. All such experts are likely to know each other and might not want to openly disagree or share the same limited experience, especially if this disagreement might have some impact on issues not related to monitoring.
3	Ask each expert to score, in writing, enforcement for three broad categories of tobacco control measures on a scale of 1 to 3 (minimally, moderately or highly enforced: (1) smoke-free environments, (2) direct advertising, (3) indirect advertising (promotion and sponsorship).
4	Review the expert's opinion at the national level. The GTCR national focal point: review these answers and clarify any pending issue or obtain more information regarding widely different answers.
5	Review national findings at the regional level. Consistency and comparability of the national answers could then be compared at the regional level and scores revised if needed.
6	Integrate results globally.

GTCR: Global Tobacco Control Report

Table 4.4 GTCR Protocol to Assess in Country Enforcement of Smoke-Free Environments, and Direct and Indirect Advertising Bans

the pitfalls of such measures described earlier, and the data collectors are aware of some countries where there are very close links between the experts. The system, however, is successful enough to serve as a basis for the next round of data collection.

Data sources:

In most cases, the source of primary data is legislation as assessed by country level informants. Informants also have to provide supporting information for these answers in the form of legal texts and official policy guidelines, although supporting documents are generally incomplete. This information is then assessed at the regional level by a regional data collector and then again at the worldwide level. For most countries, this process results in a large flow of communications where questionnaire answers are questioned, answered again, documented, and finally validated by all. The validation process thus spreads throughout data collection and is completed by a final country validation of the data. This validation includes official signing off on the questionnaire answers by an authorized civil servant[2]. Additional primary data sources are the actual knowledge of the country informant on local policies regarding the treatment of tobacco cessation. For example, the informant has to collect information on the national availability of quitlines, as well as counselling services for cessation. This

[2] This validation process was not followed for the European region in the first release, since the source of the data was the already validated data used for the European Tobacco Control Report, in addition to minor updates.

information is not backed by supporting documents unless policy papers, or even leaflet advertisements for these services, are available.

Some questionnaire items proved difficult to respond to. The simplicity of the questionnaire could not capture well the complexity of national tax data. Government spending on tobacco control also proved an elusive piece of information, because such expenditures are not clearly labelled and are often scattered across many budget items. It is therefore likely that future editions of the GTCR will need to modify the questionnaire to better capture very complex information. Finally, it proved easier to handle prevalence data through WHO's Global InfoBase than through prevalence-related questions on the questionnaire, given the clear advantage and networks InfoBase developed over the years.

Geographical coverage:

The geographical coverage is very wide, including all 193 WHO member states; although 21 countries, mostly from the Western Pacific and Americas regions, did not participate in the first release. At this stage, the GTCR questionnaire does not collect information on subnational jurisdictions, but does ask questions to certify the existence of such measures, in order to consider the feasibility of collecting these measures in the next release.

Timeliness, frequency of data collection, and trend:

The GTCR will be released annually, even if annual differences are minimal. Some changes in the data might occur despite the absence of any new measures, since a much larger team will be in charge of assessing questionnaire answers and comparing it to legislation; hence, possible revisions and refinements. The GTCR will keep an annual record of the situation in each country, which will permit trends analysis.

Reflections on the future of tobacco control monitoring systems

None of the existing monitoring systems fully meets all the criteria developed in the second part of this section, and thus it remains difficult to answer the questions outlined in the introduction without undertaking a detailed country analysis and relying on experts' opinion (Joossens & Raw, 2006). In other words, reliable, comparable, comprehensive, and ready-to-use time series on the prevalence of tobacco use and tobacco control measures do not exist and cannot be related to each other. This means that given the current stage of existing data, it remains challenging to properly and systematically assess all aspects of tobacco control as a public policy intervention at the international level, although the GTCR offers a good basis to do so if developed properly.

A new context

The environment of tobacco control has evolved very rapidly over the past few years and many initiatives either directly promote policy monitoring systems or create a strong demand for them. A major change has been the reversal of the tide in most high-income countries, with decreasing prevalence and number of smokers. However, despite prevalence rates that are also often decreasing in low- and middle-income countries, higher demographic growth will inevitably lead to deaths on a massive scale. Tobacco companies are also instituting shifts in their operations that are geared to these new markets. For this reason, tobacco control needs to quickly implement the same shift and undertake massive efforts in low- and middle-income countries.

Many factors could help this shift. The most important factor, and one that is often forgotten, is that tobacco control is now a tried and tested policy, with a tried and tested network of dedicated individuals and institutions. Tobacco control advocates can build on a lot of existing knowledge, experience, and successes, as well as failures. Awareness is also much higher, as not even the tobacco industry can argue anymore that tobacco is not bad for health.

The WHO FCTC is also a major structuring element for tobacco control. By signing it, a country *de facto* accepts its premises and commits itself in front of the world

community to enact very specific tobacco control measures, and report on the implementation of their international treaty obligations. By virtue of being a treaty, the WHO FCTC makes tobacco control a concern that is much broader than health, but an altogether international affairs issue; hence, additional pressure through linkage with other "high politics" issues.

Finally, new private and highly significant initiatives, such as the large donation by New York City Mayor Michael Bloomberg add fuel and momentum to tobacco control. These initiatives not only help strengthen existing efforts, such as the WHO FCTC, but also help empower tobacco control advocates who can then set the standards at a higher level and convince governments to follow suit. This new focus on tobacco control is thus a fantastic opportunity to start working on monitoring systems, as it creates a new demand for such information. It is time to rethink tobacco control based on past experience and highlight some of the improvements that should be implemented. These obviously have to do with the nature and analysis of the data, but mostly with the capacity to gather them.

Capacity for relevant data collection

Tobacco control is also a field that has greatly evolved with our knowledge of tobacco and of its social determinants and impact. Secondhand smoke, for example, was not a major concern for public policy before research clearly linked it to specific health conditions (US Department of Health and Human Services, 1986). Realizing that youth prevalence is a major explanatory factor for future adult prevalence, has meant that tobacco control could adopt much more aggressive policies towards this specific market. Knowledge that some of the harm caused by tobacco to the cardiovascular system can be reversed within a few years of cessation, has given a tremendous boost to cessation policies. The tobacco industry's reaction to original advertising bans has prompted a policy reaction that now stretches to promotion and sponsorship, etc. Linking smoking further to a general discomfort and economic costs for nonsmokers, and realizing that smoking bans were also a very efficient way to help addicted smokers quit, helped justify further tobacco control in the field of secondhand smoke. The health impact on nonsmokers, however, remains a crucial underpinning for public intervention in this field.

Monitoring systems for tobacco control must thus be flexible enough to evolve and keep up with the changes in overall policy objectives, tobacco control environment, and consumption patterns. Monitoring systems for tobacco control are consequently much more than just gathering data. They involve a complex process with clear objectives and constant reassessment of policy means. The most striking implication of this policy process is the ensuing need for a dedicated network of individuals, institutions, and ongoing discussions regarding both the evolution and continuity of the system, as well as the nature and usefulness of the collected data. Health practitioners, economists, epidemiologists, data managers and collectors, government officials, and many others need a very high level of collaboration in order to set up and maintain a good tobacco control monitoring system. A prerequisite to any good monitoring system is, therefore, a good organisation, which points directly to the most important ingredients: dedicated work with regular, predictable, and stable funding.

Referring back to the questions outlined in the first paragraph of the introduction: why can't we better assess the impact of specific tobacco control policy interventions in terms of efficiency and efficacy? One important factor is the capacity to build and sustain policy monitoring systems. In fact, many initiatives were started and left incomplete, mainly because of irregular or insufficient funding (perhaps as a reflection of lack of political will). As this section made clear, a high-quality international monitoring system is first and foremost a good and stable network of competent and highly coordinated individuals and institutions. Such networks are difficult to build and maintain. In addition, close supervision of

country level activities is impossible to perform from the outside, and this necessitates close involvement of local authorities and staff, hence the absolute necessity of country buy-in.

This means that the most pressing demand from countries is in capacity building to gather and analyse data. Indeed, based on past experiences, building a sustainable tobacco control monitoring system is impossible without a prior effort to build a solid network of competent individuals and institutions, and a national level capacity that can sustain this system. Previous data collection efforts were mainly donor-driven. A network of informants was set up from various sources (ministries of health, non-governmental organisations, etc.), questionnaires were answered, stipends paid, and when funds dried up, this embryo of a network was unfortunately left to disintegrate. These data collection efforts provided highly valuable information, and individuals who worked on them were pioneers in tobacco control, but unfortunately a lot of the data cannot be used now.

The incredible opportunity that now exists, thanks to the WHO FCTC, is a global demand for capacity building, as countries will start to struggle to meet international obligations. Answering this demand quickly is crucial to build a comprehensive international network for tobacco control. This network is in turn a necessary condition to the emergence of a global tobacco control monitoring system. It follows that in this new international context, capacity building should come first with data collection undertaken as an integral part of it. This would ensure country buy-in, help keep competent data collectors in the network, and answer the needs of countries regarding the WHO FCTC. Most importantly, this would ensure that the data collection system does not vanish after a round of data collection, as it will be linked to the overall policy needs of the countries making these efforts relevant not only for international users, but also for local users. This network also needs to be expanded outside of the traditional country level individuals from ministries of health, and include officials from external affairs and economic ministries, as made possible, if not necessary, by the WHO FCTC.

Towards one effective policy data collection system

A monitoring system that is solidly anchored in a network to be assembled by a significant capacity building effort is a necessary condition for success, but surely not a sufficient condition; dispersing efforts among several systems should be avoided. Countries should not be burdened by excessive data collection, at least with regards to tobacco control. This means, for example, completing the integration of the WHO regional databases and GTCR. It also means that over the next few years, the relationship between this data collection system and the WHO FCTC should be carefully assessed. Although the WHO FCTC does not yet cover all countries and does not gather data with the aim of comparing them (at least for now), there is nevertheless a significant overlap between the COP reporting instrument and GTCR. The closer these processes are, the easier data collection becomes, and the more efficient the entire system will be.

Conclusions

This section describes the few existing data collection initiatives on policy interventions in the field of tobacco control. Only the WHO GTCR system is, at this moment, a repository of good quality information on a wide range of tobacco control policy interventions for the large majority of countries. It is also the only one with sustainable funding, and therefore the most promising initiative to support prospective national policy changes over time. Nevertheless, the GTCR only focuses on policy domains that have been proven to be effective in reducing tobacco use. Its main limitation is that it does not yet contain information about sub-national policies. All policy researchers studying policy differences between countries are encouraged to use the WHO GTCR system in their investigations.

4.2 Using production, trade, and sales data in tobacco control

Introduction

Article 20 of the Framework Convention on Tobacco Control (FCTC) calls for parties to:

"(a) establish progressively a national system for the epidemiological surveillance of tobacco consumption and related social, economic and health indicators;
(b) cooperate with competent international and regional intergovernmental organizations and other bodies, including governmental and nongovernmental agencies, in regional and global tobacco surveillance and exchange of information on the indicators specified in paragraph 3(a) of this Article" (WHO, 2003).

One can envisage that as the FCTC is progressively implemented in a substantial number of countries, a comprehensive and sustainable surveillance system will emerge. Such a system would allow advocates and researchers a one stop source of information where comparable key tobacco control statistics, such as mortality attributable to tobacco use, prevalence of tobacco use, and consumption of and trade in manufactured tobacco products are accessible. Unfortunately, such a system is not yet available. Tobacco control researchers and advocates must find important data, such as cross-country estimates of production, trade, and tobacco consumption from a variety of sources.

The objectives of this section are 3-fold: to discuss the potential usefulness of production and trade data in tobacco control, with particular attention to the advantages and disadvantages of using these data to measure tobacco consumption; to examine the use of export and import statistics for measuring the illegal cigarette trade; and to review the availability and quality of existing data.

Trade and production data in tobacco control

Data on trade and production of manufactured tobacco products can be obtained from national statistical agencies and international databases with relative ease and provide valuable information to tobacco control advocates. First, production data can provide a good indicator of the importance of the national tobacco industry at both the national and international levels and, in the absence of trade, production data can provide an accurate measure of the national tobacco market. Secondly, data on the import and export of manufactured tobacco products can provide valuable information on important, key players in the national tobacco control debate. For example, a close examination of trade patterns in tobacco products can reveal the precise origin of cigarette imports; similarly, it can identify key export markets. Such information can be invaluable in identifying important players in the national tobacco control arena. Finally, production figures can be combined with import and export figures, to provide a measure of national consumption of manufactured tobacco products that may be useful in attempting to quantify the magnitude of the smuggling market. Sales data, based on tax records, can also be used as an estimate of the consumption of various tobacco products.

Using aggregate data to measure cigarette consumption: advantages and disadvantages

Estimates of consumption and prevalence of use of tobacco products can originate from various types of data. They can be based on (self-reported) tobacco use prevalence surveys, which provide information on the proportion of tobacco users in a given population.

Prevalence data combined with tobacco use intensity data (e.g. number of cigarette smoked per day) can also yield total con-sumption estimates. Consump-tion can also be derived from aggre-gate production and trade statistics. Production plus imports minus exports will yield "apparent" consumption estimates. For example:

- cigarette consumption = cigarette production + cigarette imports – cigarette exports
- per capita cigarette consumption = cigarette consumption / (pop. 15+)

National cigarette sales data, based on sales or tax records, can also be an estimator of consumption (Guindon & Boisclair, 2003).

Prevalence surveys can provide important insights into patterns of and changes in consumption according to sex, age, income, and education (Warner, 1977). They also allow distinguishing between a change in the number of smokers and changes in consumption per smoker. On the other hand, consumption data (the number of cigarettes consumed) based on surveys can suffer from significant underreporting (Warner, 1978; Jackson & Beaglehole, 1985; Hatziandreu *et al.*, 1989; Foss *et al.*, 1998). Surveys generally provide valid estimates of prevalence (Velicer *et al.*, 1992; Patrick *et al.*, 1994; Caraballo *et al.*, 2001; Caraballo *et al.*, 2004), suggesting that the number of cigarettes smoked each day is underreported. In addition, many population-based surveys do not interview people in the military, prison, and psychiatric institutions and thus will not assess use in populations with fairly substantial smoking prevalence. Another potential limitation is the infrequent availability of trend data. Finally, the subjective nature of surveys and differences in survey methodology (questions, definitions, languages, etc.) also make comparison of estimates across countries difficult.

Aggregate production and trade statistics are objective data that eliminate the underreporting problem inherent in data based on subjective survey responses (Warner, 1977). These data are also readily available across time and countries. This feature, as well as the availability of centralized data sources using common methodologies, allows for good comparability. However, most of these large-scale tobacco statistics are only available for manufactured cigarettes. Data from the Global Youth Tobacco Survey (GYTS) indicate that more than 10% of students used tobacco products other than cigarettes, with the rate being highest in the southeast Asia region and the eastern Mediterranean region (Warren *et al.*, 2006). Specific examples include: India where tobacco consumption is dominated by use of non-cigarette tobacco (bidis, leaf tobacco etc.), resulting in cigarette consumption repre-senting only 15% of total tobacco consumption (Rijo, 2005); and Thailand where high levels of use of hand-rolled tobacco have been reported (Sarntisart, 2003)

The major problem with aggregate data is perhaps that, unlike prevalence survey-based data, they cannot be used for analyzing changes in sex, age, income, and education distribution, and they do not permit a distinction between a change in the number of smokers and changes in consumption per smoker (Warner, 1977). Other important problems include illicit trade in cigarettes and illegal manufacturing and counterfeit trade, resulting in export and import data not being registered in official figures, which may lead to under or overestimating consumption of tobacco products (WHO, 1998a). The problem of stockpiling may also emerge, as not all cigarettes will be consumed in the year they are produced or imported. If this stockpiling is significant it may bias consumption estimates. However, it is doubtful that stockpiling will affect trends since it is not likely to vary from year-to–year, although tobacco companies have been known to time cigarette stockpiling against health measures so that they appear less effective (WHO, 1998a). Transient populations will affect aggregate trade and production statistics to a varying degree. Finally, the question of measurement units can yield diverging trends and biased point estimates. More specifically:

- "Apparent" consumption will underestimate true consumption in countries where tobacco products are illegally imported and consumed, while it will overestimate true consumption where tobacco products are illegally exported to another country.
- Trade and production data can be reported in weight or in physical units. In countries where cigarette weights have not remained constant over time, cigarette consumption expressed in units and in weight can show diverging trends. For example, Australian cigarettes became progressively lighter in the late 1980s. When expressed in grams per capita, cigarette consumption in Australia fell by 4.9% between 1986 and 1990, while it increased by 5% when expressed in units (Chapman, 1992).
- Trade and production statistics for an individual country can also be reported in different units. For example, manufactured cigarette imports and exports are often reported in metric tons, while production is expressed in units. When this is the case, it is usually assumed in the calculations that one cigarette weighs one gram. But this assumption may not hold and thus bias consumption estimates. The direction of the bias will depend on two factors: the true "conversion factor," and the respective size of imports and exports. For example, in a country where production statistics are expressed in units, trade statistics in metric tons, and one gram of cigarette equals one cigarette, true consumption will be over-estimated if the country is a net importer of cigarettes, and underestimated it if the country is a net exporter.
- "Apparent" consumption will overestimate true consumption in countries with large transient populations (for example tourists or military), and small indigenous populations, such as Malta and the Maldives.

In addition to the measurement issues described above, production and trade figures reported by national statistical agencies may not accurately reflect true figures. There may be a time lag of three to six months between recording export and import statistics. It may also be the case that import statistics are recorded more rapidly and accurately because of more prevalent import duties (as compared to export duties). Finally, there may be reporting errors at the national level, and between the national statistical agencies, international agencies, and organisations that report cross-country statistics.

Production data can be used at the global level as a proxy for world consumption. It will be a poor proxy for consumption in most countries, but as world exports must equal world imports, aggregating cigarette production for all countries would do away with the problems associated with smuggling and attenuate the problems associated with measurement units. Unfortunately, because of unequal data availability through time, adding all production data points in a particular year can lead to underestimation.

Sales data based on tax records are also aggregate data, and similarly present the same general advantages and disadvantages as those described above for production and trade statistics. It should be noted, however, that sales data are not as readily available across countries and are not available in centralised databases. On the other hand, they do not suffer from the limitations associated with measuring and reporting units or stockpiling. They also present the advantage (unlike estimates obtained from trade and production statistics) of yielding consumption estimates that exclude duty-free sales, most of which are to non-residents and are not consumed in the country. Finally, sales data may be segmented by tobacco products (e.g. cigarettes, cigars, etc.), brands and brand variant (e.g. length-type, and descriptor-type, such as "light" or "mild"), and thus yield information on market shares by individual brands, brand family, and brand variant.

Population adjustments:

Total cigarette consumption can be useful to gauge the size of a tobacco market, but it does not allow for comparison across time and across countries. To achieve the latter, total cigarette consumption or sales can be weighted by population in order to provide an indicator of individual

consumption, usually by dividing total cigarette consumption by the population aged 15 years and above. The age group 0-14 is normally omitted because of its limited contribution to tobacco use (Chapman, 1992). However, differences between countries in demographic distribution and tobacco use prevalence in the 10-20 age group can be important and diminish comparability.

The use of export and import statistics for measuring the illegal cigarette trade

The gap between global exports and global imports is often used to make estimates of the overall size of cigarette smuggling. World cigarette production is known fairly accurately, and, since there are not large numbers of cigarettes in storage because they do not keep for long, world production is very close to world consumption. Global imports should thus be close to exports, after allowing for legitimate trade usually excluded from national statistics. (These are principally imports for duty-free sales to travellers, diplomatic staff, and military establishments.)

Imports, however, have long been lower than exports to an extent that cannot be explained by legitimate duty-free sales. Even the lag time of three to six months between recording export and import statistics, cannot explain the differences between them which have been high for years. Worldwide, United States Department of Agriculture (USDA) data showed that recorded cigarette exports exceeded recorded imports by more than 300 billion each year in the period 1995-2000. The only plausible explanation for these missing cigarettes is smuggling (Joossens & Raw, 1995; Joossens & Raw, 1998).

Some cautious interpretation of these results is advisable (Merriman *et al.*, 2000). Many factors may explain a discrepancy between recorded exports and imports. An analysis of data from the United Nations Commodity Trade Statistics Database (UN Comtrade) shows large discrepancies between total reported imports and exports of many goods. However, researchers admit that cigarettes are different from other commodities, as cigarette exports consistently greatly exceed imports. It is concluded that the most reasonable explanation for the observed data is that a large and growing fraction of international trade is smuggled (Merriman *et al.,* 2000).

USDA statistics for the period 2001-2004 showed that the gap between recorded cigarette imports and exports had been reduced to around 150 billion cigarettes annually. There may be different explanations for these reductions. USDA data are not always reliable at the national or worldwide level. In 2002, the USDA magazine *Tobacco: World Markets and Trade* published data which showed that the gap between exports and imports was 276 billion cigarettes in 2001. Two years later, the same magazine released figures which showed that the gap had been reduced to 126 billion cigarettes in 2001. Caution with the analysis of USDA data is necessary.

Another explanation might be that the reduction of smuggling occurred as some major international tobacco companies have reviewed their export practices due to lawsuits. The reduction of the gap may finally be explained through the increase of illegal manufacturing and counterfeit cigarette trade, which is a growing concern in many countries. The illegal nature of their production means that they are not registered in the official export and import data.

Finally, the analysis of export and import practices can also be used to study the smuggling problem at the national level. For instance, exports from the British tobacco companies to Andorra increased from 13 million cigarettes in 1993 to 1,520 million in 1997. Taking into account that almost none of these cigarettes were legally re-exported, that Andorra only has a population of 63000, and that smokers in Andorra on the whole do not smoke British brands, it was clear that these increased exports were intended for the smuggling market (Joossens & Raw, 2002). Induced by high taxes in the early 1990s, cigarette smuggling increased substantially in Canada. Virtually all smuggled cigarettes had been previously exported from Canada. As Canada did not, and still does not, export a large amount of cigarettes, exports proved to be

an accurate indicator for smuggling (Galbraith & Kaiserman, 1997). Similarly, a significant and unlikely decrease in "apparent" cigarette consumption per capita was observed in Brazil, while "apparent" consumption was rising rapidly in Paraguay in the late 1980s and early 1990s, driven by a 16-fold increase in exports to Paraguay (Shafey *et al.*, 2002).

The aforementioned examples indicate the usefulness of examining production, trade, and consumption data to gain insights into the smuggling market. That said, other methods exist and have been used to estimate the size of national smuggling market. Tobacco consumption estimated from production and trade or sales data can be compared to estimates of consumption based on prevalence surveys while taking into account under-reporting. The United Kingdom has used this method extensively to estimate the size of the smuggling market (for more details, see HM Customs & Excise, 2001). In Thailand, individuals who reported using tobacco products during face-to-face interviews, were asked to present their tobacco package to the interviewer. An examination of the health warnings (i.e. absence of warnings or a warning in a language other than Thai) can reveal if the tobacco products are likely to have been legally purchased (Sarntisart, 2003).

Availability and quality of existing data

This section describes various cross-country sources of production and trade statistics that provide information on manufactured tobacco products, and discusses their strengths and weaknesses.

United Nations Commodity Trade Statistics Database (UN Comtrade):

The United Nations Commodity Trade Statistics Database (UN Comtrade) contains detailed import and export statistics, including manufactured cigarettes and cigars, cheroots, and cigarillos reported by statistical authorities of close to 200 countries or areas (http://unstats. un.org/unsd/comtrade/). It contains annual trade (import and export) data from 1962 to the present. UN Comtrade is considered the most comprehensive trade database available and is continuously updated. Unlike other existing data sources where only total amounts are obtainable, UN Comtrade makes available the complete trade matrix. Whenever trade data are received from the national authorities, they are standardised by the United Nations Statistics Division and then added to UN Comtrade. Despite its comprehensiveness and its online availability, UN Comtrade is rarely used by tobacco control researchers and advocates.

United Nations Statistical Division (UNSD) Industrial Commodity Production Statistics Dataset:

The current version of the UNSD Industrial Commodity Production Statistics Dataset contains the entire database of industrial commodity statistics, including manufactured cigarettes and cigars, cheroots, and cigarillos covering the period 1950-2003 (1970-2003 for manufactured cigarettes). Data for the time period 1994-2003 are available in print in the *2003 Industrial Commodity Statistics Yearbook* (United Nations Statistical Division, 2003). The data contained in this database has primarily been collected from questionnaires sent yearly to national statistical authorities. However, data have also been collected from other governmental agencies, specialised agencies, intergovernmental bodies, private institutes, and associations. The UNSD Industrial Commodity Production Statistics Dataset can be considered the most reliable and comprehensive production dataset available (http://unstats.un.org/unsd/industry/ics_ intro.asp).

Food and Agriculture Organization of the United Nations' (FAO) FAOSTAT:

The Food and Agriculture Organization of the United Nations' FAOSTAT provides access to over 3 million time-series and cross-sectional data relating to

food and agriculture from over 100 countries and areas (http://faostat.fao.org/).

The FAOSTAT TradeSTAT module contains detailed agricultural trade data, including import and export statistics for manufactured cigarettes and cigars, cheroots, and cigarillos (i.e. as a grouping). Data are obtained from national statistical and agricultural agencies and are standardised, processed, and validated by the FAO Statistics Division, whereby the national commodity classification (usually the Harmonized System) is converted to the FAO commodity classification. TradeSTAT has just recently begun providing detailed trade matrices.

United States Department of Agriculture (USDA), Foreign Agricultural Service (FAS):

- *Tobacco: World Markets and Trade* (http://www.fas.usda.gov/tobacco_arc.asp)
- *Attaché Reports* (http://www.fas.usda.gov/scriptsw/AttacheRep/default.asp)

The USDA's FAS World Market and Trade reports provide the latest data on a number of agricultural commodities, outlining the current supply, demand, and trade estimates both for the USA and for many major countries. FAS international offices provide information on production, consumption, and trade of many commodities, including manufactured cigarettes. It should be noted that the data contained in these commodity and country reports are not official USDA data, but represent estimates made by FAS Attachés. The publication *Tobacco*: *World Markets and Trade* was discontinued in September 2005, while tobacco attaché reports were discontinued in January 2006.

Data from the USDA are arguably the most widely used and cited cross-national consumption and trade statistics in tobacco control research and advocacy. The WHO Global Status Report (WHO, 1997) relies almost exclusively on data from the USDA. The much cited analysis of the impact of USA trade policy on cigarette use in Asia, utilised cigarette consumption estimates that were derived from USDA data (Chaloupka & Laixuthai, 1996). Other more recent research examples include Gilmore & McKee (2004) and Gilmore & McKee (2005).

Market research reports:

There is a plethora of reports published by market research firms on the manufactured tobacco sector. Most provide country snapshots using various market size indicators including apparent consumption, which, as mentioned earlier, is constructed from trade and production figures. These reports often present market share data by brands, brand families, and companies. Many reports offer little original information (e.g. some rely almost entirely on USDA published data). The World Cigarette Reports, published by ERC Statistics International PLC, a London-based market research organisation, provides some original statistical information, including up-to-date production and trade figures for a number of countries covered (ERC Statistics International PLC, World Cigarette Markets; http://www.erc-world. com).

United Nations Population Division (UNOP) – World population prospects:

This dataset provides the official United Nations population estimates and projections pre-pared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (http://www.un. org/esa/population/publications/WPP2004/wpp2004.htm). Detailed population estimates stratified by sex and age for close to 200 countries and areas are available.

In addition to the data sources discussed above, there exists a number of initiatives that report cross-country data for smaller groupings of countries often on a regional basis. Examples include the Organization for Economic Cooperation and Development (OECD) Health Data which reports tobacco consumption estimates for OECD member states. The latest version of the OECD database was released in June 2006, and contains a number of comparable statistics on health and health systems across OECD countries. The database contains more than

1200 series covering a wide range of health topics (i.e. health status, health care resources, health care utilisation, expenditure on health, health care financing, social protection, pharmaceutical market, and non-medical determinants of health). OECD Health Data is developed jointly by the OECD Secretariat and the Institut de Recherche et d'Étude en Économie de la Santé (IRDES), a French research institute specialising in health economics and health statistics. The data are compiled from national statistical agencies and other relevant national organisations (http://www.oecd.org/document/30/0,2340,en_2825_495642_12968734_1_1_1_1,00.html).

A second cross-country data source is the Interstate Statistical Committee of the Commonwealth of Independent States (CIS), Official Statistics of the Countries of the CIS (the CIS is comprised of Azerbaijan, Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Uzbekistan, and Ukraine). The CIS database (http://www.cisstat.com/eng/cd-offst.htm) is updated annually and contains annual data on more than 3500 socioeconomic indicators from 1980 for all CIS countries. Another data source is the Asian Development Bank (ADB) Key Indicators (http://www.adb.org/Documents/Books/Key_Indicators/2006/default.asp), which reports up-to-date manufactured cigarette statistics for a number of countries. Most data, but not all, contained in the OECD, CIS, and ADB databases are also available in the UN databases discussed earlier. However, these databases offer a relatively easy opportunity to compare estimates of consumption and production from multiple sources.

Discussion

It is important to point out that a large amount of the data published and available from the data sources described above can differ substantially. In particular, the trade data reported by the USDA, UN Comtrade, and the FAO differ widely at times. This makes it important to use the best available data by first comparing data from multiple sources.

It is generally the opinion that data from UN Comtrade (export and import) and UNSD (production) are the most reliable and comprehensive available. FAO's TradeSTAT is a good source of data that can be used alongside UN Comtrade. Of particular concern are the country data published by the USDA. They are often significantly at odds from those published by other organisations, such as the United Nations Statistical Division and the FAO, or by national statistical agencies. For a great number of low- and middle-income countries (e.g. Albania, Algeria, Bangladesh, Bolivia, Ecuador, Jordan, Lebanon, and Viet Nam), USDA cigarette production and trade data appear at best to be an extrapolation based on a "guesstimate." As discussed earlier, an examination of what is often referred to as the size of the smuggling market (the difference between total exports and total imports) yields a very different picture if looking at data from the USDA or FAO (UN Comtrade does not publish global figures of manufactured cigarettes import and export) (Guindon & Boisclair, 2003). For these reasons, it is strongly suggested to use published USDA data for low- and middle-income countries with great caution.

Researchers and advocates interested in production, trade, and consumption estimates from a single country are advised to always look first at potential local and national primary sources of information, such as government statistics agencies and ministries of trade and industry.

4.3 Data sources for monitoring global trends in tobacco use behaviours

Introduction

The purpose of this section is to describe the data collection efforts for global surveillance on tobacco use in youth and adults. We include only those surveillance systems that are cross-national and on-going. The youth surveys are school-based with a target survey population of students between 11 and 15 years of age, the primary age of smoking initiation in many countries. The surveillance systems described in this section include: The European School Survey Project on Alcohol and Other Drugs (ESPAD) (ESPAD, 2007), the Global School-Based Student Health Survey (GSHS) (GSHS, 2007), the Global Youth Tobacco Survey (GYTS) (GYTS, 2007), and the Health Behavior in School-Aged Children Survey (HBSC) (HBSC, 2007). The adult surveys have been population-based and target a wider age range (in most cases aged 15-64 or age 18+) than the youth surveys. The adult surveillance systems described include: the Global Adult Tobacco Survey (GATS) (GATS, 2007), the International Tobacco Control Survey (ITC) (ITC, 2007), and the STEPwise Approach to Chronic Disease Factor Surveillance (STEPS) (STEPS, 2007). A description of these youth and adult surveillance systems will be provided in regards to purpose, methodology, survey instrument, survey administration procedures, data analyses, dissemination of information, and utility in monitoring and evaluating articles from the WHO FCTC (WHO, 2003).

Youth

Purpose

European School Survey Project on Alcohol and Other Drugs (ESPAD):

The Pompidou Group is a multi-disciplinary cooperation forum to prevent drug abuse and illicit trafficking in drugs, set up in 1971 and incorporated into the Council of Europe in 1980. At that time, the group recognized the need for countries to collect data on alcohol, tobacco, and other drug use as it relates to public health policy and programmes (ESPAD, 2007). Three points were apparent:

1) Systematic information is generally best gathered through surveys
2) Large-scale, on-going surveys have been conducted, but only in a few countries and not as part of a cross-nationally coordinated system
3) Previous surveys had different methodologies and content, so cross-country comparisons were not possible.

To address these data gaps, the Pompidou Group developed a standard questionnaire for school-based surveys which was pilot tested in eight European countries. Further work was not done until the early 1990s, when the Swedish Government convened a meeting of 21 European countries to build on the work of the Pompidou Group by developing a system for simultaneously collecting school-based data using a common methodology. This resulted in the development of the ESPAD project which has now completed four cycles of data collection: 1995, 1999, 2003, and 2007. Future expansion will occur on a four year cycle. The countries that have participated in ESPAD are shown in Table 4.5.

The goal of ESPAD is to collect cross-nationally comparable data on alcohol, tobacco, and other drug use among students in European countries, and monitor the trends in alcohol and drug use. This is very important as it relates to the European Union (EU) action plan on drugs (EPHA, 2007) and the WHO Europe declaration about young people and alcohol (WHO, 2007b).

1995	1999	2003	2007
Croatia	Croatia	Croatia	Croatia
Cyprus	Cyprus	Cyprus	Cyprus
Czech Republic	Czech Republic	Czech Republic	Czech Republic
Denmark	Denmark	Denmark	Denmark
Estonia	Estonia	Estonia	Estonia
Faroe Islands	Faroe Islands	Faroe Islands	Faroe Islands
Finland	Finland	Finland	Finland
Hungary	Hungary	Hungary	Hungary
Iceland	Iceland	Iceland	Iceland
Italy	Italy	Italy	Italy
Latvia	Latvia	Latvia	Latvia
Lithuania	Lithuania	Lithuania	Lithuania
Malta	Malta	Malta	Malta
Norway	Norway	Norway	Norway
Poland	Poland	Poland	Poland
Portugal	Portugal	Portugal	Portugal
Slovakia	Slovakia	Slovakia	Slovakia
Slovenia	Slovenia	Slovenia	Slovenia
Sweden	Sweden	Sweden	Sweden
Turkey	Ukraine	Turkey	Turkey
Ukraine	United Kingdom	Ukraine	Ukraine
United Kingdom	Greece	United Kingdom	United Kingdom
	Greenland	Greece	Greece
	Bulgaria	Greenland	Greenland
	France	Bulgaria	Bulgaria
	FYR Macedonia	France	France
	Netherlands	Netherlands	FYR Macedonia
	Romania	Romania	Netherlands
	Russian Federation	Russian Federation	Romania
		Austria	Russian Federation
		Belgium	Austria
		Isle of Man	Belgium
		Germany	Isle of Man
		Switzerland	Germany
			Switzerland
			Serbia
			Monaco
			Armenia
			Bosnia & Herzegovina

Table 4.5 Countries Participating in the European School Survey Project on Alcohol and Other Drugs (ESPAD) by Year of Completion

Global School-Based Student Health Survey (GSHS):

The GSHS was developed by WHO (Health Promotion Division) in collaboration with UNAIDS, UNESCO, and UNICEF, and with technical assistance from the United States Centers for Disease Control and Prevention (CDC), Division of Adolescent and School Health in 2001. A school-based survey, GSHS is designed to help countries assess behavioural risk and protective factors among students aged 13-15 years. GSHS data can be used by countries to develop priorities, establish programmes, and advocate for resources for school and youth health programmes and policies. It also can be used by international agencies, countries, and others to make comparisons across countries regarding the prevalence of health behaviours and protective factors and to analyze trends in the behaviours. Implementation of GSHS started in 2003; by the end of 2006, 23 countries had completed a GSHS (Table 4.6).

Global Youth Tobacco Survey (GYTS):

In 1998, WHO and the CDC convened a meeting in Geneva to address the issue of data needs on tobacco use among youth across all Member States of WHO. Three summary points were made at this meeting:

1) Research from developed countries has found that the majority of smokers begin using tobacco products well before the age of 18 years (Perry *et al.*, 1994; Kessler, 1995)
2) Little information exists on tobacco use among youth in developing countries
3) To bridge this data gap and to promote tobacco control for all WHO Member States, WHO's Tobacco Free Initiative (TFI) and CDC's Office on Smoking and Health (OSH) agreed to support the development of the GYTS (GTSS Collaborating Group, 2005).

Implementation of GYTS started in 1999 with 12 countries (Table 4.7). By the end of 2006, 150 countries had conducted the GYTS, and over 50 countries had repeated the survey at least one time. In 2007, 11 countries conducted GYTS for the first time, 46 completed a second round, and 8 a third round.

Health Behavior of School-aged Children Survey (HBSC):

In 1982, the HBSC was initiated by researchers from England, France, and Norway. The purpose of HBSC is to collect data on young people's health and well-being, health behaviours, and the social context in which youth live. Data from HBSC have been used to influence health promotion and education policy at national and international levels. In the mid-1980s, HBSC was adopted by the WHO European Regional Office as a WHO collaborative study. HBSC was developed by a multi-disciplinary network of researchers from countries in Europe and North America. It was first conducted in 1983/84 (5 countries), then in 1985/86 (13 countries), and then every four years: 1989/90 (16

2003	2004	2005	2006	2007
China	Chile	Botswana	Egypt	Cayman Islands
Kenya	Guyana	Lebanon	Guatemala	Djibouti
Philippines	Jordan	Oman	Morocco	**Philippines**
Swaziland	Namibia	Senegal	Tanzania	India
Uganda	Zambia	Tajikistan	Uruguay	Libya
Venezuela		United Arab Emirates		Peru
Zimbabwe				St Lucia
				St Vincent & Grenadines

Table 4.6 Countries Participating in the Global School-Based Student Health Survey (GSHS) by Year of Completion

1999	2000	2001	2002	2003	2004	2005	2006	2007
Barbados	Antigua & Barbuda	American Virgin Islands	Bahrain	Argentina	Afghanistan	American Samoa	Algeria	Bangladesh
China	Argentina	British Virgin Islands	Barbados	Belize	Albania	Cyprus	Angola	Barbados
Costa Rica	Bahamas	Burkina Faso	Botswana	Benin	American Virgin Islands	Egypt	Burkina Faso	Bosnia & Herzegovina
Fiji	Bolivia	Colombia	Brazil	Bolivia	Antigua & Barbuda	Fiji	Czech Republic	Botswana
Jordan	Chile	Cuba	Bulgaria	Bosnia & Herzegovina	Armenia	Gaza Strip/ West Bank	DRC*	Brazil
Poland	Dominica	Ecuador	Costa Rica	Cambodia	Bahamas	Ghana	Eritrea	Bulgaria
Russian Federation	Ghana	Egypt	Czech Republic	Chile	Bangladesh	Haiti	Guatemala	Cambodia
South Africa	Grenada	Gaza Strip/ West Bank	Guam	Cook Islands	Belarus	Iraq	India	Colombia
Sri Lanka	Guyana	Haiti	Guatemala	Cote d'Ivoire	Bhutan	Kuwait	Indonesia	Cook Islands
Ukraine	India	Kenya	Latvia	Croatia	Cuba	Lebanon	Jamaica	Costa Rica
Venezuela	Indonesia	Kuwait	Lesotho	Djibouti	Dominica	Lithuania	Micronesia	Croatia
Zimbabwe	Jamaica	Lebanon	Mozambique	El Salvador	Dominican Rep	Macao	Morocco	Djibouti
	Mariana Islands	Lithuania	Oman	Estonia	Greece	Malawi	New Zealand	Ecuador
	Mexico	Macao	Panama	Ethiopia	Grenada	Mauritania	PNG*	Estonia
	Micronesia	Malawi	Senegal	FYR Macedonia	Guyana	Niger	Samoa	FYR Macedonia
	Montserrat	Mali	Seychelles	Georgia	Kazakhstan	Palau	Taiwan, China	Georgia
	Palau	Mauritania	South Africa	Honduras	Kosovo	Sudan	Timor-Leste	Hungary
	Peru	Morocco	St Kitts & Nevis	Hungary	Kyrgyzstan	Swaziland	Tuvalu	Iran
	Philippines	Myanmar	Syria	Iran	Malaysia	Thailand	Uruguay	Jordan
	Singapore	Nepal	Togo	Jordan	Moldova	Ukraine	USA	Kenya
	Suriname	Niger	Uganda	Laos	Montenegro	United Arab Emirates	Vanuatu	Laos
	Trinidad & Tobago	Nigeria	United Arab Emirates	Libya	Namibia			Latvia
	USA	Saudi Arabia	USA	Maldives	Philippines			Lesotho
		St Lucia	Zambia	Mariana Islands	Puerto Rico			Libya
		St Vincent & Grenadines		Mauritius	Qatar			Maldives
		Sudan		Mongolia	Rep of Korea			Mali
		Swaziland		Nicaragua	Romania			Mexico
		Tunisia		Pakistan	Russian Federation			Mongolia
		Uruguay		Paraguay	Somalia			Mozambique
				Peru	Suriname			Myanmar

Table 4.7 Countries Participating in the Global Youth Tobacco Survey (GYTS) by Year Survey Was Completed

1999	2000	2001	2002	2003	2004	2005	2006	2007
				Poland Serbia Slovakia Slovenia Sri Lanka Tanzania Turkey Venezuela Viet Nam Yemen Zimbabwe	Taiwan, China Tajikistan USA			Nepal Oman Panama Peru Philippines Qatar Saudi Arabia Senegal Serbia Seychelles Slovakia Slovenia Somalia South Africa Sri Lanka St Kitts & Nevis St Lucia St Vincent & Grenadines Syria Togo Trinidad & Tobago Tunisia Turkey Uganda Venezuela Viet Nam Yemen Zambia Zimbabwe

Table 4.7 Countries Participating in the Global Youth Tobacco Survey (GYTS) by Year Survey Was Completed

*DRC = Democratic Republic of the Congo; PNG = Papua New Guinea

1983/84	1985/86	1989/90	1993/94	1997/98	2001/02	2005/06
Austria	Austria	Austria	Austria	Austria	Austria	Austria
Denmark*	Denmark*	Denmark*	Denmark	Denmark	Denmark	Denmark
England	Finland	Finland	Finland	England	England	England
Finland	Norway	Norway	Norway	Finland	Finland	Finland
Norway	Belgium	Belgium	Belgium	Norway	Norway	Norway
	Hungary	Hungary	Hungary	Belgium	Belgium	Belgium
	Israel	Scotland	Israel	Hungary	Hungary	Hungary
	Scotland	Spain	Scotland	Israel	Israel	Israel
	Spain	Sweden	Spain	Scotland	Scotland	Scotland
	Sweden	Switzerland	Sweden	Spain	Spain	Spain
	Switzerland	Wales	Switzerland	Sweden	Sweden	Sweden
	Wales	Netherlands*	Wales	Switzerland	Switzerland	Switzerland
	Netherlands*	Canada	Netherlands	Wales	Wales	Wales
		Latvia	Canada		Netherlands	Netherlands
		N Ireland	Latvia	Canada	Canada	Canada
		Poland	N Ireland	Latvia	Latvia	Latvia
			Poland	N Ireland	N Ireland	N Ireland
			Czech Rep	Poland	Poland	Poland
			Estonia	Czech Republic	Czech Republic	Czech Republic
			France	Estonia	Estonia	Estonia
			Germany	France	France	France
			Greenland	Germany	Germany	Germany
			Lithuania	Greenland	Greenland	Greenland
			Russia	Lithuania	Lithuania	Lithuania
			Slovakia	Russia	Russia	Russia
				Slovakia	Slovakia	Slovakia
				Greece	Greece	Greece
				Portugal	Portugal	Portugal
				Rep of Ireland	Rep of Ireland	Rep of Ireland
				USA	USA	USA
					FYR Macedonia	FYR Macedonia
					Italy	Italy
					Croatia	Croatia
					Malta	Malta
					Slovenia	Slovenia
					Ukraine	Ukraine
						Iceland
						Luxembourg
						Romania
						Turkey

Table 4.8 Countries Participating in the Health Behaviour in School-Aged Survey (HBSC) by Year of Completion

*Survey conducted after schedule

countries), 1993/94 (25 countries), 1997/98 (29 countries), 2001/02 (36 countries), and 2005/06 (40 countries) (Table 4.8; http:// www.hbsc.org/countries.html).

Methodology

European School Survey Project on Alcohol and Other Drugs (ESPAD):

The ESPAD is a school-based survey with the target population being students who are, or will be, 16 years old during the year the data are collected. ESPAD follows a cluster sample design to produce nationally representative data; but the sampling can be either total population sampling, simple cluster sampling, two-stage cluster sampling, or stratified cluster sampling. A minimum of 2400 completed interviews are recommended by ESPAD. If students aged 15-16 are in two or more grades, the survey protocol recommends that all these grades should be included in the sampling frame.

Global School-Based Student Health Survey (GSHS):

A school-based survey, GSHS is conducted primarily among students aged 13-15 years. It uses the same methodology as GYTS (discussed below in the GYTS methodology section). In 11 countries, GYTS and GSHS are currently being conducted simultaneously, sharing sampled schools, but different classes are randomly selected for each survey.

Global Youth Tobacco Survey (GYTS):

The GYTS is a school-based survey of a defined geographic area that can be a country, a province, a city, or any other geographic entity (Centers for Disease Control and Prevention, 2001). Samples are selected as follows:

- The country research coordinator identifies the grades that correspond to students aged 13-15 years in the educational system.
- The research coordinator prepares a database of schools that include the identified grades. Each school is assigned a unique identifier to facilitate school selection. The number of students enrolled in each school grade to be surveyed is added to the database, which forms the survey sampling frame. The amount of work involved in creating this database varies from country to country. In some countries, the creation of the sampling frame has been the most time consuming part of the GYTS.
- The database is sent to the CDC, where the GYTS sample is drawn using a two-stage cluster sample design. Schools are selected with probability proportional to school enrolment size during the first stage, and then classes within participating schools are selected as a systematic equal probability sample with a random start during the second stage. All students in the selected classes are eligible to participate in the survey. For this two-stage sample design, statistical analysis conducted by the CDC (Centers for Disease Control and Prevention, 1999b) has found that, for most sample designs, a minimum of 1500 completed student interviews is needed to obtain a precision level of ± 5% for a given estimate. WHO and CDC use this information to work with the countries to determine the sample size of schools and students needed for each site. The desired sample size is then adjusted for anticipated non-response at the school, class, and student levels. Sample size is further increased if regional or population subgroup estimates are requested within the country.

Since classes are carefully identified to correspond to students 13-15 years old, the majority of selected students are in this age group. However, all students in the selected classes are eligible to participate regardless of age; therefore, some students were younger than 13 years or older than 15 years.

Health Behavior in School-Aged Children Survey (HBSC):

The HBSC is a school-based survey with the target population of students 11, 13, and 15 years old. The desired mean age for the three age groups is 11.5, 13.5, and 15.5 respectively. In some countries, each age group can be found in the same school year, while in others they may be found across years with a proportion of

students being advanced or held back. Cluster sampling is used where the primary sampling unit is school class. The survey is carried out as a nationally representative sample in each participating country. The recommended sample size for each of the three age groups is set at approximately 1500 students. This target population assumes a 95% confidence interval of ± 3% around a proportion of 50% and a design effect of 1.2, based on analysis of existing HBSC data.

Given differences in school systems, age at admission, and the degree of advancement and holding back among students, imposing a uniform approach is problematic in the HBSC. To overcome this complexity, age has been a priority for sampling, with students of the relevant age selected across school years. This position can be further complicated when the target population is split across different levels of schooling, such as primary and secondary. Where the number of classes eligible for sampling is unknown, probability proportionate to size sampling is used, making use of actual or estimated school size. In some countries, to minimize the number of participating schools, classes for one age group were randomly sampled in schools, and then classes drawn from other grades in the same schools. In order to produce mean ages of 11.5, 13.5, and 15.5, the survey is administered at appropriate times of the year.

Survey Instrument

European School Survey Project on Alcohol and Other Drugs (ESPAD):

Questions on alcohol, tobacco, and drugs are included in the ESPAD. There are core questions that all countries are encouraged to include, as well as optional and module questions that may be added. Countries are encouraged to field-test their questionnaire. The final version of the questionnaire is translated into each language needed within country then back-translated into English as a quality control check. The research protocol specifies that questionnaires should be administered anonymously.

Tobacco-related questions in ESPAD include: lifetime cigarette use, use of cigarettes in the last 30 days (i.e. current cigarette smoking), age of initiation of cigarette smoking, number of friends who smoke cigarettes, and number of siblings who smoke.

Global School-Based Student Health Survey (GSHS):

The GSHS includes questions on alcohol, and other drug use; dietary behaviours; hygiene; mental health; physical activity; protective factors; respondent demographics; sexual behaviour; tobacco use; and violence and unintentional injury. Each country develops their questionnaire, which can include core modules, core-expanded questions, and country-specific questions. The final questionnaire is self-administered in classes during one regular class period. The questions are translated into the appropriate language of instruction for the students and pilot tested for comprehension. All questions share common characteristics to enhance the flow of the survey and comprehension by the student.

Core GSHS questions on tobacco use include: age of initiation, cigarette smoking during the past 30 days (i.e. current cigarette smoking), use of other tobacco products during the past 30 days, attempts to stop smoking during the past 12 months, exposure to secondhand smoke during the past 7 days, and use of tobacco by parents or guardians.

Global Youth Tobacco Survey (GYTS):

The GYTS questionnaire is a self-administered, school-based instrument consisting of a core set of questions that are used by all countries, unless the information is not relevant in that country (e.g. pro-cigarette advertising is not permitted in Singapore, so these questions are omitted). In addition, there is an optional set of questions from which a country can draw depending on its needs and priorities. Specific guidelines are followed for questionnaire translation into local languages and pilot testing. The final questionnaire is the responsibility of each participating country.

The 2007 core GYTS questionnaire consists of 54 questions, and includes items on the following topics: prevalence of tobacco use, age of initiation,

exposure to tobacco advertising, perceptions and attitudes on behavioural norms with regard to tobacco use among young people, media and advertising, school curriculum, and secondhand smoke exposure. The GYTS core questionnaire includes information that can be used to monitor seven Articles of the WHO FCTC (Articles 8, 12, 13, 14, 16, 20, and 21) (WHO, 2003).

Health Behavior in School-Aged Children Survey (HBSC):

The HBSC questionnaire consists of a mandatory set of items that each country is required to include: health and well-being, tobacco smoking, alcohol use, cannabis use, physical activity, sedentary behaviour, eating habits, body image, weight control, body weight, oral health, bulling, physical fighting and victimization, and injuries. Countries can also include items specific to their national needs. The final questionnaire includes items on health and health-related behaviours and the life circumstances of young people.

HBSC questions on tobacco use include: lifetime tobacco use, current tobacco smoking, rate of consumption of cigarettes, and age of initiation of daily smoking

Survey administration procedures

European School Survey Project on Alcohol and Other Drugs (ESPAD):

The ESPAD recommends data collection during March/April, and the research protocol states that the survey should be conducted during a week that does not proceed a holiday. Schools that cannot perform the survey during an assigned week are encouraged to use the following week. When possible, the survey should be conducted at the same time in all classes in a school; thus, avoiding the possibility of discussion among students in the school. Each ESPAD researcher decides who to use for survey administration (i.e. teachers, research assistants). ESPAD provides the survey administrator with written instructions on how to conduct the data collection in a class.

Global School-Based Student Health Survey (GSHS):

A survey coordinator in each country manages the GSHS. The coordinator is responsible for the overall management of the project, and functions as a liaison with other agencies and organisations in the country, as well as with WHO and CDC. Survey coordinators are trained during regional workshops on the specific procedures to follow for data collection and data management.

Global Youth Tobacco Survey (GYTS):

As with GSHS, the GYTS is managed by a survey coordinator in each country. Regional training workshops are held each year to train the coordinators on data collection and data management procedures. The intent is to standardize the data collection and management procedures across the countries and within each country across time. A GYTS research manual was developed, which includes detailed procedures for administering the GYTS in schools. The manual is modified for each subsequent GYTS training to meet the specific needs of the countries in those trainings. The manual includes information on obtaining school participation, procedures for completing all survey forms, protocol in the classroom, and instructions for returning the completed forms to CDC for data processing. The GYTS uses a generic answer sheet, which allows for a maximum of 99 questions, with eight response categories available per question. There are no open ended questions, skip patterns, or multiple response questions in the GYTS. The completed answer sheets are scanned through an optical reader. Edits for consistency and out-of-range responses are performed for each question. Data quality issues of this type have been rare; consistency failures or out-of-range responses rarely exceed 5% per question.

The GYTS is administered during one class period. GYTS administration procedures were designed to protect students' privacy by assuring that student participation was anonymous and voluntary. Before the survey is administered, each country follows local procedures for obtaining parental permission and institutional review.

Health Behavior in School-Aged Children Survey (HBSC):

In most cases, data collection for HBSC is between October and May. Data collection consists of the delivery of questionnaires to selected schools for teacher administration. In some schools, researchers administer the survey in the classes in an attempt to minimize teacher burden. Once collected, the data are sent to the HBSC Internal Data Bank at the Norwegian Social Science Data Services for cleaning and final country dataset preparation.

Data analysis

Global School-Based Student Health Survey (GSHS) and Global Youth Tobacco Survey (GYTS):

Both GSHS and GYTS data are weighted to adjust for sample selection (school and class levels), non-response (school, class, and student levels), and post-stratification of the sample population relative to the grade and sex distribution in the total population. The weighting factor consists of the inverse of the probability of selection for each school; the inverse of the probability of selection of each classroom; within each selected school, a school level; non-response adjustment calculated by school enrolment size category (small, medium, large); school non-response calculated within each tertile; a class level, non-response adjustment factor calculated for each school; a student level, non-response adjustment factor calculated by class; and a post-stratification adjustment factor calculated by sex and grade. The computer program SUDAAN (http://-www.rti.org/SUDAAN/) is used to compute standard errors, 95% confidence intervals, and weighted prevalence estimates.

Health Behavior in School-Aged Children Survey (HBSC):

HBSC employs a clustered sampling design, where the primary sampling unit is the class (or school) rather than the individual student, as in a simple random sample. Given such a design, the students' responses cannot be assumed to be independent, as students within the same class or school are more likely to be similar to each other than to students in general. Cluster sampling, therefore, results in standard errors that tend to be higher than would be the case if the same size of sample were obtained using a simple random sample. Consequently, standard errors must be calculated using an appropriate method that takes into account the correlation of young people in schools or classes (SUDAAN, STATA (http://www.stata.com/), and EPI INFO (http:// www.cdc.gov/epiinfo/) are statistical packages developed for the analysis of complex survey data). In addition, a number of countries and regions stratify their samples, classifying the sample frame into smaller units (i.e. geographical areas) to ensure coverage of all regions. This stratification is likely to reduce standard errors and should be taken into account when they are being calculated.

Dissemination of Information

Information on the ESPAD can be found at http://www.espad.org. In addition, cross-national reports for study years 1995, 1999 and 2003 are available from the Swedish Council for Information on Alcohol and Other Drugs.

Information on the GSHS can be found at http://www.who.int/chp/gshs/en and http://www.cdc.gov/gshs. Country datasets can be obtained on both websites.

Information on the GYTS can be found at http://www. cdc.gov/tobacco/global. The GYTS website includes Country Fact Sheets, Country GYTS Reports, and access to country datasets. In addition, over 45 articles using GYTS data have been published in peer reviewed journals, such as *Lancet, Tobacco Control, and Morbidity and Mortality Weekly Reports*.

Information on the HBSC can be found at http://www.hbsc.org. Over 160 articles have been published featuring HBSC data, including recent articles in the *European Journal of Public Health, Health Education, and the Journal of Adolescent Health*.

2007	2008
Bangladesh	China
Brazil	Indonesia
Egypt	Mexico
India	Pakistan
Russian Federation	Philippine
Thailand	Poland
	Turkey
	Ukraine
	Viet Nam

Table 4.9 Countries Participating in the Global Adult Tobacco Surveys (GATS) by Year of Survey Completion

Summary

Comparison of youth survey content

All four surveys measure tobacco use prevalence (See Table 4.12 for a full comparison of measures by survey). ESPAD and HBSC ask only about cigarette smoking. GSHS and GYTS ask about cigarette smoking, as well as use of other tobacco products. All four surveys ask about age of initiation of cigarette smoking, however ESPAD, GSHS, and GYTS ask about first use, whereas HBSC asks about initiation of daily smoking.

ESPAD, GSHS, and GYTS ask respondents about secondhand smoke exposure, but use different indicators to assess exposure. ESPAD and GYTS ask about number of friends who smoke and ESPAD asks about number of siblings who smoke. GSHS and GYTS ask about exposure to secondhand smoke at home and in public places during the week prior to the survey, as well as smoking behaviour of parents.

GYTS assesses school curriculum by asking students if they were taught about the dangers of smoking in the year prior to the survey, if they discussed reasons why people their age smoke, and if they were taught about the specific health effects of smoking. The other three surveys do not include items to assess school curriculum components.

GYTS measures exposure to pro-tobacco media messages by asking students if they have seen actors smoking in movies, videos, or on TV; if they saw ads on billboards or in newspapers for tobacco products; and if they have an object with a cigarette brand logo on it. GYTS also asks students if they have seen anti-tobacco media messages. The other three surveys do not include indicators of media exposure to tobacco advertising.

GSHS and GYTS ask students about cessation behaviour. Both surveys ask students if they have tried to quit smoking in the year prior to the survey. GYTS also

2002	2003	2004	2005	2006	2007
Australia	Australia	Australia	Australia	Australia	Australia
Canada	Canada	Canada	Canada	Canada	Canada
United Kingdom	United Kingdom	Ireland	Ireland	China	China
United States	United States	United Kingdom	Malaysia	Ireland	France
		United States	Republic of Korea	Mexico	Germany
			Scotland	Scotland	Ireland
			Thailand	United Kingdom	Malaysia
			United Kingdom	United States	Mexico
			United States	Uruguay	New Zealand
					Scotland
					Thailand
					United Kingdom
					United States

Table 4.10 Countries Participating in the International Tobacco Control Survey (ITC) by Year of Survey Completion

asks students if they received help to quit smoking and from whom, and measures tobacco dependency using a standard indicator of addiction (time to first cigarette). The other two surveys do not include measures of cessation.

GYTS assesses minors' access to tobacco products by asking current smokers where they usually get their cigarettes, if they have been refused purchase of cigarettes when they tried to buy them in a store, and if they have been offered free cigarettes by a tobacco company representative. The other three surveys do not include measures of minors' access to tobacco products.

Limitations of youth survey content

There are several limitations inherent in each of the youth surveys. First, the target populations are young people in school, and by definition, school-based surveys do not attempt to collect information about the portion of the youth population that is out of school. School-based surveys are thus not representative of the entire youth population in any country. The extent to which the information collected by a school-based survey is not representative of the total youth population varies by country. Second, the school-based surveys described in this section conduct anonymous and self-administered interviews giving each student in a selected class one chance to participate. Students who miss class or refuse to participate are not represented in the sample. Third, extensive reliability testing of all the instruments used by the different surveys has not been completed; however, questions on tobacco use in GYTS also appearing in the CDC's Youth Risk Behavioral Survey (YRBS), have been shown to have good test-retest reliability in a study conducted in the USA (Brener *et al*., 1995).

Adults

Purpose

Global Adult Tobacco Survey (GATS);

In 2006, the GATS was initiated with funds from the Bloomberg Foundation to reduce tobacco use in low- and middle-income countries. The initiative places a priority on countries with the greatest number of smokers. More than half of the world's smokers live in fifteen countries: China, India, Indonesia, Russia, Bangladesh, Brazil, Mexico, Turkey, Pakistan, Egypt, Ukraine, Philippines, Thailand, Viet Nam, and Poland (Table 4.9).

In addition to the CDC Foundation, other key partners in the Bloomberg Initiative include the Campaign for Tobacco Free-Kids, the World Lung Foundation, the Johns Hopkins Bloomberg School of Public Health, and the WHO. Partners are charged with working collaboratively to promote international support for tobacco control policies, increase effective advocacy, and implement tobacco-free programming. Specifically, the partner organisations will:

- Refine and optimize tobacco control programmes to help smokers stop and prevent children from starting
- Support public sector efforts to pass and enforce key laws and implement effective policies, in particular, to tax cigarettes, prevent smuggling, change the image of tobacco, and protect workers from exposure to other people's smoke
- Support advocates' efforts to educate communities about the harms of tobacco and to enhance tobacco control activities so as to help make the world tobacco-free
- Develop a rigorous system to monitor the status of global tobacco use.

The CDC Foundation worked with partners around the world, particularly with the WHO, and in high-burden countries, to develop GATS (i.e. establish systematic, standardised global surveillance and monitoring of the tobacco epidemic).

International Tobacco Control Survey (ITC):

The ITC Project began in 2002 as a prospective cohort study tracking and comparing the impact of national-level tobacco policies among representative samples of adult smokers in four countries: the USA, Canada, the United Kingdom, and Australia (Table 4.10). In 2004, ITC was expanded to include smokers from Ireland and a new cohort of smokers from the UK, to evaluate the 2004

2002	2003	2004	2005	2006	2007
Ethiopia	Algeria	American Samoa	Burundi	Aruba	Angola
Fiji	Bangladesh	Cook Islands	Cote d'Ivoire	Iran	Barbados
Oman	Cameroon	Jordan	DRC*	Kuwait	Botswana
Samoa	**India**	Lebanon	DPRK*	Mauritania	Cambodia
	Indonesia	Maldives	Egypt	Mongolia	Cape Verde
	Kenya	Myanmar	Iraq	**Sri Lanka**	China
	Marshall Islands	Nauru	Kiribati	Thailand	Cuba
	Micronesia	Pakistan	Mauritius	Vanuatu	Curacao
	Palau		Mozambique	Zambia	Dominica
	Sri Lanka		Nepal		Dominican Rep
	Syria		Saudi Arabia		Equatorial Guinea
			Solomon Islands		Gaza Strip
			Tokelau		Ghana
			Tuvalu		Grenada
			Zimbabwe		**Kenya**
					India
					Iran
					Laos
					Libya
					Paraguay
					PNG*
					St Kitts & Nevis
					South Africa
					Tanzania
					Trinidad & Tobago
					Togo
					Turks & Caicos
					Uganda
					Uruguay
					Viet Nam
					Zimbabwe

Table 4.11 Countries Participating in the WHO STEPwise Approach to Surveillance (STEPS) by Year of Survey Completion

*DRC = Democratic Republic of the Congo; DPRK = Democratic People's Republic of Korea; PNG = Papua New Guinea

Ireland smoke-free policy. In 2005, the collection of ITC countries was further expanded to include cohorts of smokers in Malaysia, Republic of Korea, Scotland, and Thailand. In 2006, ITC was further expanded to include China, Mexico, and Uruguay; in 2007 France, Germany and New Zealand joined on. The objective of the ITC is to apply rigorous research methods to evaluate the psychosocial and behavioural effects of national level tobacco control policies. The ITC Project uses multiple country controls, longitudinal designs, and theory-driven mediational models that allow tests of hypotheses about the anticipated effects of given policies.

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):

In 2000, the 53rd World Health Assembly passed a resolution in support of the need to prevent and control non-communicable diseases (NCD). The goal of the resolution was to support WHO Member States in their efforts to reduce morbidity, disability, and premature mortality related to NCDs. Development of a NCD surveillance system was one of the primary objectives of this effort, and WHO STEPwise approach to Surveillance (STEPS) was developed to meet this need. The WHO STEPS is a simple, standardised method for collecting, analyzing, and disseminating data in WHO member countries.

By using the same standardised questions and protocols, all countries can use STEPS information not only for monitoring within-country trends, but also for making comparisons across countries. The approach encourages the collection of small amounts of useful information on a regular and continuing basis.

As a surveillance system, STEPS provides information on NCD risk behaviours that countries can use for better public health policy decision-making. The goal of STEPS is to build the capacity of countries to develop and maintain an integrated, systematic, data collection system that collects data on NCDs and their risk factors, including information on tobacco use (specific tobacco questions included in STEPS are discussed later in the section). There are currently two primary STEPS surveillance systems: the STEPwise approach to risk factor surveillance, and the STEPwise approach to stroke surveillance. The survey is currently being implemented in over 80 countries with new countries coming on board on a regular basis (Table 4.11). STEPS is active in all WHO regions except EURO (where existing surveillance systems are already in place for NCD risk factors). Nearly all AFRO countries have done or plan to do STEPS surveys.

Survey methodology

Global Adult Tobacco Survey (GATS):

The GATS is a household survey of adults aged 15-64 years. The sample domains include complete population coverage, except for areas that have special country circumstances (e.g. conflict areas, remote areas). In addition, institutional populations (e.g. prisons, dormitories, hospitals) are excluded. A multi-stage sampling design was used to include all household members aged 15-64 from a sample of households, with one individual randomly selected per household. Interviews were completed face-to-face. In this survey, a probability sample is required; therefore, an appropriate method of random sampling is used in each sampling stage so that selection probabilities can be determined for all sampling units in each stage, and the probability of selection for each respondent (computed as the product of stage-specific probabilities) is known. Aside from needed oversampling (e.g. by urban/rural and region), random selection was used in each stage in a way that makes selection probabilities among respondents as equal as possible. Substitution or replacement sampling was not allowed in any stage of the sample design. Four stages are included in the sample design: primary sample units (PSU) of the smaller, or the smallest, recognized geopolitical area units with current statistical population (i.e. individual or household); count data and quality cartographic maps (e.g. county, census tract, or block group, rather than state in the USA); secondary sampling units (SSU) of recognized geopolitical subunits to the area units used for PSUs; individual housing/dwelling units (see Census website for definitions of these geographic terms), or small groups (<10) of neighboring housing units (HUs compact segments); and finally, within-household sampling of one study-eligible household resident from a roster of residents 15-64 years of age.

Targeted sample sizes (for both genders combined) for urban and rural respondents should be approximately 4000 each. This can be accomplished by selecting the same number of PSUs in urban and rural strata. These estimates were arrived at by specifying the following parameters in the sample size calculation needed to detect differences in key rates (smoking prevalence) between survey rounds: 95% confidence error margins of ≤ 3 percentage points

for tobacco use estimates of 40% at any given round, 80% power (Type I error of 0.05 and two-sided test alternatives), and a design effect of 2.0 (arising out of the effect of cluster sampling). Samples in each round are independently chosen and should be proportionate for all demographic categories except level of urbanization and region.

International Tobacco Control Survey (ITC):

The ITC Project stratifies the country population into several geographic regions. Quotas were assigned for the number of respondents (age 18 and over) in each of the strata, in order to ensure representation proportional to a measure of regional population size. In the original four countries, eligible households were then selected by random digit dialing methods until the within-stratum quotas were met. As the survey was expanded to countries that had less complete phone coverage, the ITC employed multistage cluster sampling across entire countries (Thailand and Malaysia), or within key geographical areas (Mexico, Uruguay, China), which was implemented with face-to-face interviews. A household was deemed eligible for inclusion in the survey if it contained at least one eligible smoker. In households with multiple eligible smokers, the Next Birthday method was used to select a single respondent. No substitution within the household was permitted, except where it was known that the selected respondent would be absent for the entire fieldwork procedure.

A ten minute recruitment survey was first conducted to screen for eligibility. A thank you letter and financial compensation were mailed immediately after the recruitment calls. In order to avoid call-scheduling bias, recruitment calls were conducted at various times of the day and on different days of the week. If a respondent agreed to participate, but did not keep a main survey appointment, up to 25 attempts to follow-up were made at varying times of day. In addition, respondents could complete the main survey during two or more calls if requested (Thompson *et al.*, 2006).

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):

The STEPS is an adult survey conducted with a sample of 25-64 year olds (although many countries survey young adults age 15-18) in the household setting (Steps 1 and 2) and in the clinic setting (Step 3). Five different sample designs are supported. In general, the sample is a multi-stage cluster sample of at least 2000 adults.

Prior to survey implementation, STEPS completes preliminary phases including: defining target population, sample size, sampling frame and design, selecting sample participants, and documenting sample selection. Its methodology emphasizes sampling a target population that at a minimum comprises adults aged 25-64; wider age ranges are permissible, but not narrower.

The sampling scheme to be followed will depend on the size of the population, geographic area to be sampled, and available resources. Stratification of the population to be sampled is often done according to physical location of the sampling units (e.g. urban versus rural). Proportional or disproportional allocation of sampling units per strata may be enforced. Simple random sampling or multi-stage cluster sampling are followed, and both can be utilized in conjunction with stratification. STEPS recommends the use of multi-stage cluster sampling when conducting national surveys.

For the actual drawing of the sample, sampling probability proportional to size is used. Once the selection of the household and/or individuals is completed, data collection begins through interviewing. If collection of clinical data is planned, participating clinics are identified and clinical registration, blood collection, and biochemical measurement forms are compiled along with biological samples.

Survey Instrument

Global Adult Tobacco Survey (GATS):

The core GATS questionnaire was developed by an expert committee, including representatives from WHO (regional and country offices), CDC, and international tobacco control experts. The core instrument was tested in cognitive laboratory procedures in March 2007, and was piloted tested in the

Philippines and India in April and May 2007. Results from the cognitive laboratory and pilot studies were used to finalize the core questionnaire at a meeting of the expert committee in June 2007.

The core GATS questionnaire includes indicators on tobacco prevalence (smoking and smokeless tobacco use), exposure to secondhand smoke, cessation, risk perceptions, knowledge and attitudes, exposure to media, and price and taxation issues.

International Tobacco Control Survey (ITC):

The ITC questionnaire was developed by a multidisciplinary team of international tobacco control experts. A pilot survey, including the screener and main survey, was conducted among 125 respondents; the instrument was revised as a result of the pretest. The questionnaire has been revised at each subsequent wave, but the core of the instrument has remained essentially the same to facilitate comparisons and modeling over time. Apart from minor variations in colloquial language, the same questionnaire was used in all four English speaking countries; translations are used in the other countries.

Due to the objectives of ITC and the eligibility requirements of respondents, the questionnaire includes questions on a wide range of tobacco-related behaviours, knowledge, and attitudes that are targeted to current smokers. These indicators include: daily cigarette consumption, weekly cigarette consumption, type of product smoked (hand-rolled or manufactured; menthol, Virginia, or blended; pieces in pack; filtered or non-filtered), cigarette brand preference, duration of smoking (time since respondent started smoking), dependency (time to first smoke), current use of tobacco products other than cigarettes (including cigars, pipes, chewing tobacco, snuff, and other products), number of closest friends who smoke, smoking policy at respondent's place of work, support for smoking regulations in indoor public areas, knowledge of health effects and diseases caused by smoking, beliefs about dangers of different tobacco products, perception of relative danger of tobacco products other than cigarettes, a module of questions regarding warning labels, a module of questions about pro-tobacco advertising, a module of questions about awareness campaigns that shows dangers of smoking or encourages quitting, desire to quit, number of quit attempts, duration of last smoke-free period, knowledge and use of cessation support products, cessation services available (doctor or health professional, telephone quit line, clinics, participation in international events such as Quit and Win Contests), and perceived difficulty to quit smoking.

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):

The STEPS instrument covers three different levels of "steps" of risk factor assessment through the collection of the following types of data:

- Questionnaire
- Physical measurements
- Biochemical measurements

Step 1 gathers information on risk factors that can be obtained from the general population by questionnaire. This includes information on socio-demographic features, tobacco use, alcohol consumption, physical inactivity, and fruit/vegetable intake. Step 2 includes objective data by simple physical measurements needed to examine risk factors that are physiologic attributes of the human body, such as height, weight, waist circumference (for obesity), and blood pressure. Step 3 carries the objective measurements of physiologic attributes one step further with the inclusion of blood samples for measuring lipid and glucose levels.

The STEPS tobacco questions were drawn from WHO's Guidelines for Controlling and Monitoring the Tobacco Epidemic (WHO, 1998a). Core tobacco use questions include: current smoking of any tobacco products (such as cigarettes, cigars, or pipes); current daily smoking of tobacco products; age of initiation of daily smoking; and daily consumption of tobacco (manufactured cigarettes, hand-rolled cigarettes, pipes full of tobacco, cigars/cheroots/cigaril-los, or other). Expanded tobacco use questions include: ever smoke daily; age when stopped smoking daily; current use of smokeless tobacco, such as snuff, chewing tobacco, betel; current daily use of

smokeless tobacco products; number of times a day use smokeless tobacco products; and ever daily use of smokeless tobacco, such as snuff, chewing tobacco, betel. Currently no data are collected on cessation, secondhand smoke exposure, exposure to pro-tobacco media and advertising, economics, knowledge, and attitudes.

Survey administration procedure

Global Adult Tobacco Survey (GATS):
Survey administration of GATS consists of a coordinated effort between WHO (regional and country offices), CDC, and the country GATS coordinator working in the Ministry of Health. Each country GATS coordinator identifies possible companies or agencies that can carry out the survey. WHO and CDC meet within country with the GATS coordinator to make the final selection, and follow-up on all details with the company chosen, including timeline, budget, training of interviewers, and other tasks as relevant.

GATS interviews are conducted in households by trained interviewers. Survey teams are used, which consist of a supervisor and interviewers. The supervisor has the responsibility of leading the team, identifying the correct geographic location for the selection of the households, assigning interviewers to houses, and conducting quality control checks on each interviewer. The interviewers conduct the interview with the appropriate person in each household and maintain high quality standards.

International Tobacco Control Survey (ITC):
Survey administration for ITC has been handled by contracting companies. Waves 1 and 2 of the survey were conducted in Canada and the USA by Environics Research group. Waves 1 and 2 in Australia and the UK, and all countries that participated in Waves 3 and 4, were conducted by Roy Morgan Research. Senior representatives of the companies participated in the protocol design, in order to ensure standardization of the survey administration and calling protocol across survey sites. All calling specifications, final questionnaires, and daily reports were reviewed and monitored by the ITC Research Team, at the University of Waterloo, to maintain consistency across survey firms and countries.

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):
WHO conducts Regional STEPS Training Workshops for country STEPS research coordinators. Part 3 in the WHO STEPS Surveillance Manual (available at http://www.who.int/chp/steps/manual/en/index.html) includes a "Training Guide" for how to plan, prepare for, and deliver training to the data collection, data entry, and data analysis teams. STEPS has three separate trainings: interviewer training, data entry training, and data analysis training. Part 4 in the WHO STEPS Surveillance Manual covers details regarding data collection, data entry and data management, and data analysis.

Data analysis

Global Adult Tobacco Survey (GATS):
GATS data are weighted to adjust for sample selection, non-response, and post-stratification of the sample population. Since it uses a multistage sample design, estimates of standard errors must be adjusted to take into account the design effect. Specific statistical analysis products are required that can accommodate the complex weighting considerations. The computer program SUDAAN was used to compute standard errors, 95% confidence intervals, and weighted estimates.

International Tobacco Control Survey (ITC):
The ITC sampling design was chosen to provide a random, unbiased, representative sample of adult smokers within each geographic stratum. In order to adjust for disproportionate selection and under-coverage of population subgroups, weights in Wave 1 were calculated for each respondent to adjust for number of residential phone lines and adult smokers in the household. These weights were adjusted to produce recruitment weights, so that estimates of total numbers of smokers in age-sex groups agreed with current smoking prevalence numbers in the country. The weights were also

adjusted for attrition between the recruitment and the main survey. In subsequent waves, weights were created for longitudinal or cohort analyses for respondents who completed two or more waves. Cross-sectional weights were calculated to incorporate newly recruited respondents

The ITC uses a complex survey design; therefore, standard error estimators need to be adjusted to take into account the design effect. Specific statistical analysis products are required that can accommodate the complex weighting considerations. These packages include SUDAAN, WesVar (http://www.westat.com/wesvar/), STATA, and SAS (http://www.sas.com).

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):

Part 4, Section 3 of the WHO STEPS Surveillance Manual includes discussion of the tasks that are needed to analyze STEPS data. STEPS recommends the country data analysts work with a survey statistician for advice and support (if none is available then the country coordinator can receive assistance from the STEPS team in Geneva). They also suggest that the country coordinator use EPI INFO (version 3.3 or higher), or other similar statistical software packages, for data analysis. STEPS provides technical support and training for EPI INFO, and training for analysts for data cleaning, weighting, and analysis, upon request. The STEPS sampling workbook contains spreadsheets for calculating weights. STEPS assists the country coordinator in producing a Fact Sheet showing key findings from the survey, which can be used for quick dissemination of the results.

Dissemination of Information

Global Adult Tobacco Survey (GATS):

Dissemination of GATS information is a primary focus of WHO and CDC. A website for easy access to the GATS data, reports, and country Fact Sheets is being developed by WHO (this site should be available by the end of 2008).

International Tobacco Control Survey (ITC):

Publications by ITC researchers, and other authors, featuring the ITC data can be found at http://www.itcproject.org.

STEPwise Approach to Chronic Disease Factor Surveillance (STEPS):

Part 4, Section 4 of the WHO STEPS Surveillance Manual includes information on reporting and disseminating STEPS results. Countries are encouraged to disseminate the results from their survey in a timely manner after survey completion. The results can help:

1) Raise awareness about preventing chronic disease and their risk factors
2) Guide public health policy and interventions to address chronic diseases
3) Assist and inform future health research

STEPS encourages the coordinators to prepare Fact Sheets and Country Reports. The Fact Sheet should be a short summary of the key results and used for immediate dissemination. The Country Report should be comprehensive and include: the overall rationale, scope of the survey, the sampling design, details of the methods for data collection, detailed results of the survey, and implications for future health and planning. It should be widely distributed to relevant government agencies and sponsoring organisations, non-governmental organisations that could use the information, public, government and institutional libraries, press and other media outlets, and websites. Detailed information about STEPS can be found at http://www.who.int/chp/steps/en/.

Summary

Comparison of adult survey content

All three surveys measure tobacco use prevalence and consumption levels of various products (see Table 4.13 for a full comparison of measures in each survey). GATS, ITC, and STEPS ask about cigarette smoking and use of tobacco products other than cigarettes. All four surveys ask about age of initiation of daily cigarette smoking; however, GATS asks about first use of cigarettes for smokers who are not daily smokers.

GATS and ITC query respondents about secondhand smoke exposure, but use different indicators to assess exposure. Both ask about smoking policies in respondents' homes and workplaces. ITC asks about number of closest friends who smoke. GATS has questions about the rules concerning smoking in respondents' homes and if there are other members of the household that smoke.

GATS and ITC assess respondents' knowledge and beliefs about the health effects of smoking. Both surveys include a battery of questions about the relationship between smoking and a variety of diseases and conditions. GATS and ITC ask respondents about their perceptions of the relative danger of tobacco products other than cigarettes.

GATS and ITC include questions about exposure to pro- and antitobacco media messages. Both surveys ask about respondents' exposure to pro-tobacco advertising in a variety of media, such as billboards, point of sale, radio, television, and movies. Respondents' viewing of health warning labels on tobacco packaging is also asked about in both surveys. GATS and ITC also include questions on price and taxation.

GATS and ITC ask respondents about cessation behaviour. Both surveys ask respondents about their motivation to quit smoking; unsuccessful quit attempts; and knowledge of cessation support products, such as nicotine replacement therapies and antidepressants. A measure of tobacco dependency, using a standard indicator of addiction (time to first cigarette) and respondents' perceived difficulty to quit smoking, is applied in both surveys.

ITC includes questions that assess policy-specific mediators and psychosocial mediators of policy impact.

Limitations of adult survey content

There are some limitations of ITC, GATS, and STEPS. The longitudinal design of ITC is intended to evaluate the impact of policies on smokers. The sampling methodology screens households for smokers as the target population, although participating countries have the option of including an additional sample of non-smokers. The ITC samples are designed to be representative of the smoker population of the countries or of major geographic areas within the countries; they are not designed to assess national or regional levels of tobacco prevalence. The primary limitation facing GATS at this time is the question of coverage and sustainability. The Bloomberg Foundation intends to fund future expansion and repetition of GATS, but whether this funding can lead to expansion of GATS to all WHO Member States and include provisions for repeat rounds, is unknown. As a multirisk survey, STEPS has limits on the number of tobacco-related questions that can be included. Further, STEPS is dependent on countries to follow statistically valid protocols for sample design and field procedures. It also has limited quality control measures in place to assure compliance with the protocols.

Discussion

Public health surveillance involves "…the ongoing systematic collection, analysis, and interpretation of outcome specific data for use in planning, implementation, and evaluation of public health practice" (Taylor & Bettcher, 2000). As of March 14, 2007, 145 of the 192 WHO Member States had ratified WHO FCTC. An important feature of the WHO FCTC is the call for countries to establish programmes for national, regional, and global surveillance as stated in Article 20:

Research, surveillance and exchange of information – "The Parties shall establish, as appropriate, programmes for national, regional, and global surveillance of the magnitude, patterns, determinants and consequences of tobacco consumption and exposure to tobacco smoke. Towards this end, the Parties should integrate tobacco surveillance programmes into national, regional, and global health surveillance programmes so that data are comparable and can be analyzed at the regional and international levels, as appropriate" (WHO, 2003).

One of the primary goals of the WHO FCTC is the development, implementation, and evaluation of effective tobacco control pro-

WHO FCTC Article	European School Survey Project on Alcohol and Other Drugs (ESPAD)	Global School Health Survey (GSHS)	Global Youth Tobacc Survey (GYTS)	Health Behaviour of School-aged Children (HBSC)
Article 20: **Research, surveillance and exchange of information.** The Parties shall establish, as appropriate, programmes for national, regional, and global surveillance of the magnitude, patterns, determinants, and consequences of tobacco consumption and exposure to tobacco smoke. Towards this end, the Parties should integrate tobacco surveillance programmes into national, regional, and global health surveillance programmes so that data are comparable and can be analysed at the regional and international levels, as appropriate.	In the early 1990s, the Swedish Government convened a meeting of 21 European countries to build on the work of the Pompidou Group by developing a system for simultaneously collecting school-based data using a common methodology. This resulted in the development of the ESPAD project which has now completed three cycles of data collection: 1995, 1999, and 2003. Future expansion of ESPAD will occur on a four year cycle.	GSHS data can be used by countries to develop priorities, establish programmes, and advocate for resources for school health and youth health programmes and policies. GSHS also can be used by international agencies, countries, and others to make comparisons across countries regarding the prevalence of health behaviours and protective factors and to analyze trends in the behaviours. Implementation of GSHS started in 2003; by the end of 2006, 24 countries had completed a GSHS.	Initiated in 1999, GYTS was developed by WHO Headquarters, WHO Regional Offices, and CDC. By the end of 2006, 150 countries had completed at least one round of GYTS; of these, 44 countries have completed a second round. In 2007, 17 countries conducted the survey for the first time, 31 countries were prepared to conduct a second round, and 42 trained to conduct the survey in the future.	Data from HBSC has been used to influence health promotion and health education policy at national and international levels. In the mid-1980s, HBSC was adopted by the WHO European Regional Office as a WHO collaborative study. HBSC was developed by a multidisciplinary network of researchers from countries in Europe and the United States. HBSC was first conducted in 1983/84 (5 countries), in 1985/86 (13 countries), and then every four years: 1989/90 (16 countries), 1993/94 (26 countries), 1997/98 (29 countries), 2001/02 (36 countries), and 2005/06 (41 countries).
Prevalence Article 21: **Reporting and exchange of information.** Each Party shall submit to the Conference of the Parties, through the Secretariat, periodic reports on its implementation of this Convention, which should include the following: information on surveillance and research as specified in Article 20 (Research, surveillance, and exchange of information)	- Lifetime cigarette use - Use of cigarettes in the last 30 days (i.e. current cigarette smoking) - Age of initiation of cigarette smoking	- Age of initiation - Cigarette smoking during the past 30 days (i.e. current cigarette smoking) - Use of other tobacco products during the past 30 days	- Lifetime cigarette use - Initiated smoking before age 10 - Cigarette smoking during the past 30 days (i.e. current cigarette smoking) - Current use of tobacco other than cigarettes - Never smokers susceptible to initiate smoking in the next year	- Lifetime tobacco smoke - Current tobacco smoking - Consumption of cigarettes - Age of initiation of daily smoking

Table 4.12 European School Survey Project on Alcohol and Other Drugs (ESPAD), Global School-Based Student Health Survey (GSHS), Global Youth Tobacco Survey (GYTS), and Health Behavior of School-Aged Children (HBSC) Measures That Can Be Used to Monitor the WHO Framework Convention for Tobacco Control (FCTC)

WHO FCTC Article	ESPAD	GSHS	GYTS	HBSC
Exposure to Second-hand Smoke Article 8: **Protection from exposure to tobacco smoke.** Each Party shall adopt and implement in areas of existing national jurisdiction, as determined by national law, and actively promote at other jurisdictional levels the adoption and implementation of effective legislative, executive, administrative, and/or other measures, providing for protection from exposure to tobacco smoke in indoor workplaces, public transport, indoor public places, and, as appropriate, other public places.	- Number of friends who smoke cigarettes - Number of siblings who smoke	- Exposure to second-hand smoke during the past 7 days - Use of tobacco by parents or guardians -	- Exposed to smoke from others in their home - Exposed to smoke from others in public places - Think smoking should be banned from public places - Use of tobacco by parents	
School Article 12: **Education, communication, training and public awareness.** Each Party shall promote and strengthen public awareness of tobacco control issues, using all available communication tools, as appropriate. Towards this end, each Party shall adopt and implement effective legislative, executive, administrative, or other measures, to promote public awareness of, and access to, information regarding the adverse health,economics, and environmental consequences of tobacco production and consumption.			- During past year in school, students were taught about dangers of smoking - During past year in school, students discussed reasons people their age smoke - During past year in school, students were taught about the effects of smoking	

Table 4.12 European School Survey Project on Alcohol and Other Drugs (ESPAD), Global School-Based Student Health Survey (GSHS), Global Youth Tobacco Survey (GYTS), and Health Behavior of School-Aged Children (HBSC) Measures That Can Be Used to Monitor the WHO FCTC

WHO FCTC Article	ESPAD	GSHS	GYTS	HBSC
Media and Advertising Article 13: **Tobacco advertising, promotion, and sponsorship.** Parties recognize that a comprehensive ban on advertising, promotion, and sponsorship would reduce the consumption of tobacco products			- During the past month, saw actors smoking on TV, in videos, or in movies - During the past month, saw ads for cigarettes on billboards - During the past month, saw ads for cigarettes in newspapers or magazines - During the past month, saw ads for cigarettes at sporting events, fairs, concerts, or community events - Have an object with a cigarette brand logo on it	
Cessation Article 14: **Demand reduction measures concerning tobacco dependence and cessation.** Each Party shall develop and disseminate appropriate, comprehensive, and integrated guidelines based on scientific evidence and "best practices," taking into account national circumstances and priorities, and shall take effective measures to promote cessation of tobacco use and adequate treatment for tobacco dependence		- Attempts to stop smoking during the past 12 months	- Current smokers who desire to stop smoking - Current smokers who tried to stop smoking during the past year - Current smokers who ever received help or advice from a programme or professional to help them stop smoking - Current smokers who have or feel like having a cigarette first thing in the morning	

Table 4.12 European School Survey Project on Alcohol and Other Drugs (ESPAD), Global School-Based Student Health Survey (GSHS), Global Youth Tobacco Survey (GYTS), and Health Behavior of School-Aged Children (HBSC) Measures That Can Be Used to Monitor the WHO FCTC

WHO FCTC Article	ESPAD	GSHS	GYTS	HBSC
Minor's Access and Availability Article 16: **Sales to and by minors.** Each Party shall adopt and implement effective legislative, executive, administrative, or other measures, at the appropriate level to prohibit the sales of tobacco products to persons under the age set by domestic law, national law, or age eighteen. Each Party shall prohibit or promote the prohibition of the distribution of free tobacco products to the public and especially minors.			- Current smokers who usually get their cigarettes by buying them in a store, in a shop, or from a street vendor - Current smokers who were not refused purchase of cigarettes because of their age - Students who were offered "free" cigarettes by a cigarette company representative	

Table 4.12 European School Survey Project on Alcohol and Other Drugs (ESPAD), Global School-Based Student Health Survey (GSHS), Global Youth Tobacco Survey (GYTS), and Health Behavior of School-Aged Children (HBSC) Measures That Can Be Used to Monitor the WHO FCTC

grammes in all WHO Member States.

How do data from the surveillance systems identified in this section assist countries in monitoring and evaluating articles from the WHO FCTC? As illustrated in Tables 4.12 and 4.13, these systems provide valuable indicators for measuring achievement of WHO FCTC articles. The WHO FCTC calls for countries to use consistent methods and procedures in their surveillance efforts. The surveys described in this section were created with the intention of providing internationally comparable data by employing research protocols with common sampling procedures, core questionnaire items, field procedures, and data management across survey sites.

The WHO FCTC also requires countries to monitor the treaty's application over time. Surveillance data that encompasses a broad range of information about tobacco use behaviour, and associated factors, are a necessary component of applied research that establishes evidence-based relationships between programme efforts and policy outcomes. In addition, the WHO FCTC contributes to strengthening the leadership capacity of the ministries of health and other state bodies responsible for tobacco control, not only in terms of public health advocacy, but also in negotiations with other sectors with respect to tobacco control. Finally, ongoing, systematic surveillance enhances the role of the nongovernmental sector by supporting civil society participation in monitoring the state of tobacco control efforts, and facilitating policy and programme development.

WHO FCTC Article	International Tobacco Control Policy Evaluation Survey (ITC)	STEPwise Approach to Chronic Disease Factor Surveillance (STEPS)	Global Adult Tobacco Survey (GATS)
Article 20: **Research, surveillance and exchange of information.** The Parties shall establish, as appropriate, programmes for national, regional, and global surveillance of the magnitude, patterns, determinants, and consequences of tobacco consumption and exposure to tobacco smoke. Towards this end, the Parties should integrate tobacco surveillance programmes into national, regional, and global health surveillance programmes so that data are comparable and can be anal-ysed at the regional and international levels, as appropriate.	The objective of the ITC is to apply rigorous research methods to evaluate the psychosocial and behavioural effects of national level tobacco control policies. The ITC Project uses multiple country controls, longitudinal designs, and theory-driven mediational models that allow tests of hypotheses about the anticipated effects of given policies.	STEPS provides information on NCD risk behaviours that countries can use for better public health policy decision-making. The goal of STEPS is to build the capacity of countries to develop and maintain an integrated, systematic data collection system that collects data on NCDs, and their risk factors. There are currently two primary STEPS surveillance systems: STEPwise approach to risk factor surveillance and the STEPwise approach to stroke surveillance.	The Bloomberg Initiative partners established the GATS in 15 high-burden countries to collect data on tobacco use prevalence (cigarette smoking and other tobacco use), exposure to secondhand smoke, cessation, risk perceptions, knowledge and attitudes, exposure to media, price, and taxation issues which are critical measures for tobacco control programme and policy development.
Prevalence Article 21: **Reporting and exchange of information.** Each Party shall submit to the Conference of the Parties, through the Secretariat, periodic reports on its implementation of this Convention, which should include the following: information on surveillance and research as specified in Article 20 (research, surveillance and exchange of information)	- Respondents are eligible to participate if they have smoked 100 cigarettes in their lifetime and currently smoke (manufactured or hand-rolled) cigarettes - Current daily smoking (manufactured or hand-rolled) - Daily cigarette consumption - Weekly cigarette consumption - Type of product smoked (hand-rolled or manufactured; menthol, Virginia, or blended; pieces in pack; filtered or non-filtered) - Cigarette brand preference - Duration of smoking (time since respondent started smoking) - Dependency (time to first smoke) - Current use of tobacco products other than cigarettes (including cigars, pipes, chewing tobacco, snuff, and other products)	- Current smoking of any tobacco products (such as cigarettes, cigars, or pipes) - Currently daily smoking of tobacco products - Age of initiation of daily smoking - Daily consumption of tobacco (manufactured cigarettes, hand-rolled cigarettes, pipes full of tobacco, cigars/cheroots/cigarillos, or other) - Expanded tobacco use questions include: ever smoke daily; age when stopped smoking daily; current use of smokeless tobacco, such as snuff, chewing tobacco, betel; current daily use of smokeless tobacco products; number of times a day use smokeless tobacco products; and ever daily use of smokeless tobacco, such as snuff, chewing tobacco, betel.	- Current smoking of any tobacco products (such as cigarettes, cigars, or pipes) - Current daily smoking of tobacco products - Age of initiation of daily smoking for daily smokers - Age of first cigarette smoked for less than daily smokers - Daily consumption of smoked and smokeless tobacco (including manufactured cigarettes, hand-rolled cigarettes, pipes full of tobacco, cigars/cheroots/cigarillos, water pipe rocks, or other smoked products, and snuff, chewing tobacco, betel, and other smokeless products)

Table 4.13 International Tobacco Control Policy Evaluation Survey (ITC), STEPwise Approach to Chronic Disease Factor Surveillance (STEPS), Global Adult Tobacco Survey (GATS) Measures That Can Be Used to Monitor the WHO Framework Convention for Tobacco Control (FCTC)

WHO FCTC Article	International Tobacco Control Policy Evaluation Survey (ITC)	STEPwise Approach to Chronic Disease Factor Surveillance (STEPS)	Global Adult Tobacco Survey (GATS)
Exposure to Secondhand Smoke			
Article 8: **Protection from exposure to tobacco smoke.** Each Party shall adopt and implement in areas of existing national jurisdiction, as determined by national law, and actively promote at other jurisdictional levels, the adoption and implementation of effective legislative, executive, administrative, and/or other measures, providing for protection from exposure to tobacco smoke in indoor work and public places, public transport, and, other public places.	- Number of closest friends who smoke - Smoking permitted in home - Smoking policies in places respondent goes often - Smoking policy at respondents' place of work - Support for smoking regulations in indoor public areas		- Number of family members who smoke - Smoking permitted in home - Smoking policies in places respondent goes often - Smoking policy at respondents' place of work - Support for smoking regulations in indoor public areas
Knowledge			
Article 12: **Education, communication, training and public awareness.** Each Party shall promote and strengthen public awareness of tobacco control issues, using all available communication tools, as appropriate. Towards this end, each Party shall adopt and implement effective legislative, executive, administrative, or other measures, to promote public awareness of, and access to, information regarding the adverse health, economics, and environmental consequences of tobacco production and consumption.	- Knowledge of health effects and diseases caused by smoking - Beliefs about dangers of different tobacco products - Perception of relative danger of tobacco products other than cigarettes		- Knowledge of health effects and diseases caused by smoking - Beliefs about dangers of different tobacco products - Perception of relative danger of tobacco products other than cigarettes

Table 4.13 International Tobacco Control Policy Evaluation Survey (ITC), STEPwise Approach to Chronic Disease Factor Surveillance (STEPS), Global Adult Tobacco Survey (GATS) Measures That Can Be Used to Monitor the WHO Framework Convention for Tobacco Control (FCTC)

WHO FCTC Article	International Tobacco Control Policy Evaluation Survey (ITC)	STEPwise Approach to Chronic Disease Factor Surveillance (STEPS)	Global Adult Tobacco Survey (GATS)
Media and Advertising Article 13: **Tobacco advertising, promotion, and sponsorship.** Parties recognize that a comprehensive ban on advertising, promotion, and sponsorship would reduce the consumption of tobacco products	- Module of questions regarding warning labels - Module of questions about pro-tobacco advertising on television, radio, billboards, internet, shop windows, in newspapers, restaurants, and discos - Tobacco industry sponsorship of sporting or cultural events - Module of questions about awareness campaigns that shows dangers of smoking or encourages quitting in the past 6 months		- Module of questions regarding warning labels - Module of questions about pro-tobacco advertising on television, radio, billboards, internet, shop windows, in newspapers, restaurants, and discos - Module of questions about awareness campaigns that shows dangers of smoking or encourages quitting in the past 6 months
Cessation Article 14: **Demand reduction measures concerning tobacco dependence and cessation.** Each Party shall develop and disseminate appropriate, comprehensive, and integrated guidelines based on scientific evidence and "best practices," taking into account national circumstances and priorities, and shall take effective measures to promote cessation of tobacco use and adequate treatment for tobacco dependence.	- Desire to quit - Quit attempts - Duration of last smoke-free period - Knowledge about NRT and Zyban - Use of NRT or other cessation assistants - Cessation services available (doctor or health professional, telephone quit line, clinics, participation in international events such as Quit and Win Contests) - Perceived difficulty to quit smoking		- Desire to quit - Quit attempts - Duration of last smoke-free period - Knowledge about NRT and Zyban - Use of NRT or other cessation assistants - Cessation services available (doctor or health professional, telephone quit line, clinics) - Perceived difficulty to quit smoking

Table 4.13 International Tobacco Control Policy Evaluation Survey (ITC), STEPwise Approach to Chronic Disease Factor Surveillance (STEPS), Global Adult Tobacco Survey (GATS) Measures That Can Be Used to Monitor the WHO Framework Convention for Tobacco Control (FCTC)

Summary and recommendations

The youth surveillance systems described in this section include: The European School Survey Project on Alcohol and Other Drugs (ESPAD), the Global School-Based Student Health Survey (GSHS), the Global Youth Tobacco Survey (GYTS), and the Health Behavior in School-Aged Children Survey (HBSC). The adult surveillance systems described include: the Global Adult Tobacco Survey (GATS), the International Tobacco Control Survey (ITC), and the STEPwise Approach to Chronic Disease Factor Surveillance (STEPS).

To evaluate among youth articles of the WHO FCTC, the GYTS is the only source of international data available which includes the following indicators: exposure to secondhand smoke, exposure to pro- and anti-tobacco media and advertising, cessation, minors' access to tobacco products, and school curriculum.

To evaluate among adults articles of the WHO FCTC, GATS, and ITC have the most comprehensive set of indicators, including: exposure to secondhand smoke, economics (price and taxation), cessation, product labeling, and exposure to pro- and anti-tobacco media and advertising. Where possible longitudinal studies, such as ITC, should be used for evaluating policies and programmes, because of the opportunity to examine and adjust for individual level predictors of tobacco use behaviours (see Section 2.1).

GYTS was developed, and GATS is being developed, for countries which did not have existing surveillance systems for the collection of information on tobacco use and its determinants.

Countries interested in developing a tobacco control surveillance system are encouraged to join one of these international systems. Those countries that have existing national surveys are encouraged to link to these international efforts.

5.1 Measures to assess the effectiveness of tobacco taxation

Introduction

Significant increases in cigarette and other tobacco product taxes are widely considered to be a highly effective mechanism to reduce tobacco use and, as a result, the death, disease, and economic and social costs caused by tobacco use (Jha & Chaloupka, 1999; Jha *et al*., 2006). These tax increases are effective in inducing current tobacco users to quit, preventing youth from becoming regular users, keeping former users from restarting, and reducing the amount consumed by continuing users (Chaloupka *et al.,* 2000a). When the revenues from these taxes are used to support other tobacco control efforts (e.g. enforcement of tobacco control policies, mass media information campaigns, and increased awareness of and access to cessation services and products), the impact is increased. Given this evidence, Article 6 (Figure 5.1) of the WHO FCTC, calls for Parties to the treaty to use tax and price policies to reduce tobacco use, while Article 15 (Figure 5.2) calls for the adoption and implementation of measures aimed at eliminating the illicit trade in tobacco products that can undermine the effectiveness of increased tobacco taxes.

This section focuses on measures to evaluate the effectiveness of tobacco taxation. Historically (and still the case in many countries), the primary purpose of tobacco taxation was the efficient generation of revenue for use in financing government spending. As evidence about the impact of higher taxes on tobacco use has accumulated, an increasing number of governments, particularly in high resource countries, have used higher tobacco product taxes as a tool for reducing tobacco use and its consequences (Jha & Chaloupka, 1999). Similarly, these taxes can be used to correct for the externalities caused by tobacco use, such as the health consequences of exposure to environmental tobacco smoke among non-smokers, or the financial costs of publicly financed healthcare services in treating

1. The Parties recognize that price and tax measures are an effective and important means of reducing tobacco consumption by various segments of the population, in particular young persons.
2. Without prejudice to the sovereign right of the Parties to determine and establish their taxation policies, each Party should take account of its national health objectives concerning tobacco control and adopt or maintain, as appropriate, measures which may include:
   a. Implementing tax policies and, where appropriate, price policies, on tobacco products so as to contribute to the health objectives aimed at reducing tobacco consumption; and
   b. Prohibiting or restricting, as appropriate, sales to and/or importations by international travelers of tax- and duty-free tobacco products
3. The Parties shall provide rates of taxation for tobacco products and trends in tobacco consumption in their periodic reports to the Conference of the Parties in accordance with Article 21.

WHO (2003)

Figure 5.1 WHO FCTC Article 6: *Price and Tax Measures to Reduce the Demand for Tobacco*

1. The Parties recognize that the elimination of all forms of illicit trade in tobacco products, including smuggling, illicit manufacturing and counterfeiting, and the development and implementation of related national law, in addition to subregional, regional and global agreements, are essential components of tobacco control.
2. Each Party shall adopt and implement effective legislative, executive, administrative or other measures to ensure that all unit packets and packages of tobacco products and any outside packaging of such products are marked to assist Parties in determining the origin of tobacco products, and in accordance with national law and relevant bilateral or multilateral agreements, assist Parties in determining the point of diversion and monitor, document, and control the movement of tobacco products and their legal status. In addition, each Party shall:
   a. require that unit packets and packages of tobacco products for retail and wholesale use that are sold on its domestic market carry the statement: "*Sales only allowed in (insert name of the country, subnational, regional, or federal unit*)" or carry other effective marking indicating the final destination or which would assist authorities in determining whether the product is legally for sale in the domestic market; and
   b. consider, as appropriate, developing a practical tracking and tracing regime that would further secure the distribution system and assist in the investigation of illicit trade.
3. Each Party shall require that the packaging information or marking specified in paragraph 2 of this Article shall be presented in legible form and/or appear in its principal language or languages.
4. With a view to eliminating illicit trade in tobacco products, each Party shall:
   a. Monitor and collect data on cross-border trade in tobacco products, including illicit trade, and exchange information among customs, tax and other authorities, as appropriate, and in accordance with national law and relevant applicable bilateral or multilateral agreements;
   b. enact or strengthen legislation, with appropriate penalties and remedies, against illicit trade in tobacco products, including counterfeit and contraband cigarettes;
   c. take appropriate steps to ensure that all confiscated manufacturing equipment, counterfeit and contraband cigarettes and other tobacco products are destroyed, using environmentally-friendly methods where feasible, or disposed of in accordance with national law;
   d. adopt and implement measures to monitor, document and control the storage and distribution of tobacco products held or moving under suspension of taxes or duties within its jurisdiction; and
   e. adopt measures as appropriate to enable the confiscation of proceeds derived from the illicit trade in tobacco products.
5. Information collected pursuant to subparagraphs 4(a) and 4(d) of this Article shall, as appropriate, be provided in aggregate form by the Parties in their periodic reports to the Conference of the Parties in accordance with Article 21.

WHO (2003)

Figure 5.2 WHO FCTC Article 15: *Illicit Trade in Tobacco Products*

diseases caused by tobacco. However, a number of arguments have been raised in opposition to increased tobacco taxes, including that higher taxes will promote extensive tax avoidance among continuing users, result in increased smuggling of tobacco products, unfairly burden low-income populations, and cause significant job losses. The alternative goals and potential consequences of increased tobacco taxation suggest the need to measure several outcomes resulting from a change in tobacco taxation. A simple conceptual framework for these outcomes is contained in Figure 5.3 (the bold variables are covered in the text here; the other measures are discussed elsewhere in this Handbook and will not be described in detail in this section).

There are other outcomes that can be affected by tobacco taxation, as well as by other

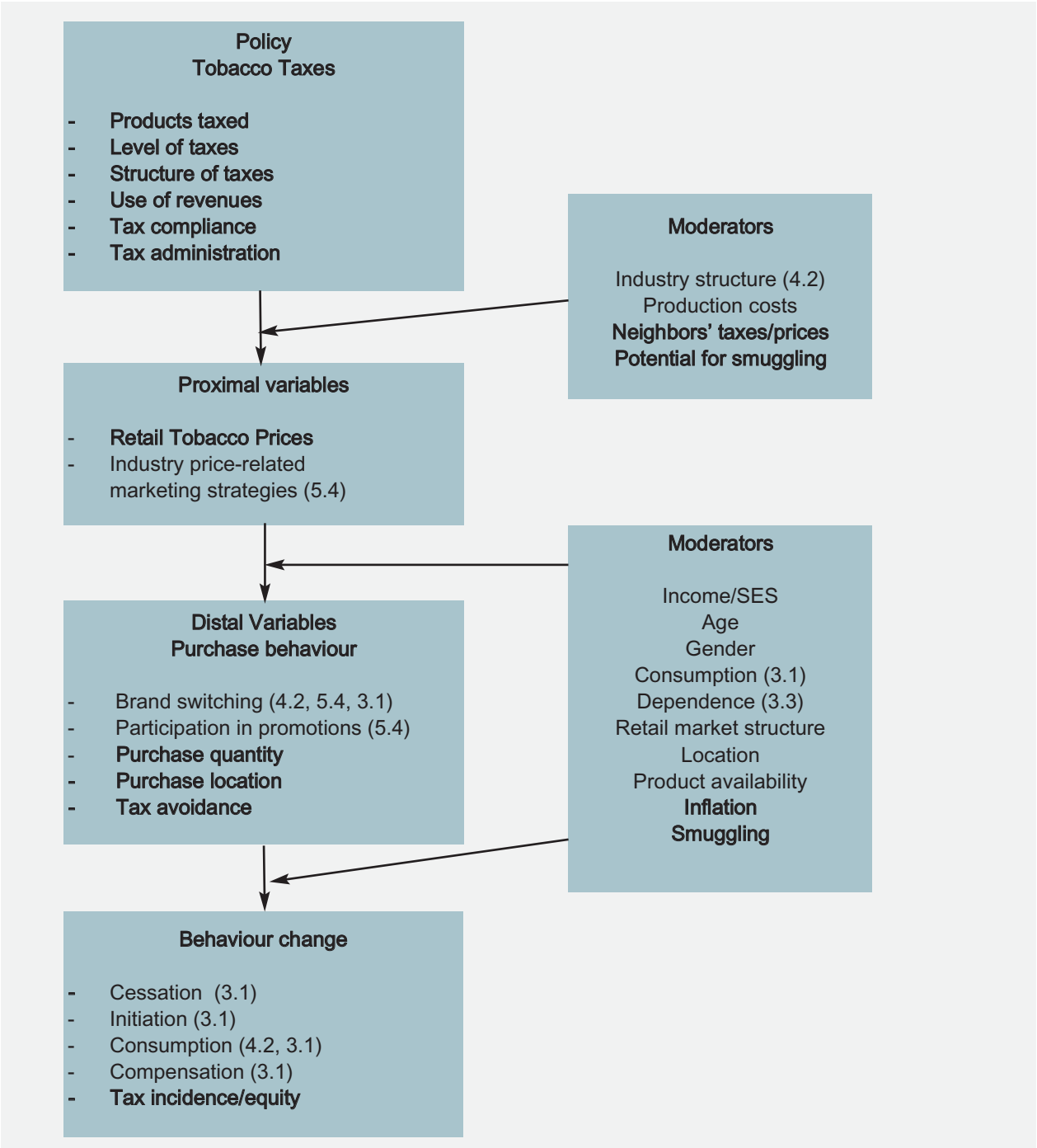


Figure 5.3 Conceptual framework for the evaluation of tobacco tax policies

Numbers in parentheses refer to sections in the Handbook covering those topics

In bold, variables covered in the main text

tobacco control policies and programmes. These include the effects of the reductions in tobacco use that result from tax increases, and other factors, on outcomes such as overall economic activity, as reflected by employment, national income, and development. Opponents of tobacco tax increases, for example, argue that higher taxes will have a negative impact as jobs in tobacco growing, manufacturing, and related activities are lost when tobacco use declines. These outcomes are beyond the scope of this Handbook; approaches to assessing these are described elsewhere (see, for example, Tool 5 of the World Bank's Economics of Tobacco Toolkit (http://www.worldbank.org/tobacco) on measuring the employment impact of tobacco control policies (Zhang, 2002)). This section will focus primarily on measuring tobacco product taxes and prices, the purchase behaviour of users, the extent of individual tax avoidance, larger scale tobacco product smuggling, and, briefly, the incidence of tobacco taxation.

Motives for tobacco taxation

It is important to understand the underlying motivation for tobacco tax increases in order to assess their effectiveness. Historically, the primary motivation for tobacco taxation was the efficient generation of government revenue, with nearly all countries having taxed tobacco products for many decades or, in some cases, centuries. Even in countries where other motives have become more important, revenue generation remains a significant factor. The less than proportionate response of tobacco product consumption to changes in tobacco product prices (relatively "inelastic demand" in the language of economists), the small number of producers, significant consumption, and lack of good substitutes make tobacco products particularly attractive targets for excise and other taxation. As Adam Smith describes in *The Wealth of Nations*, "Sugar, rum, and tobacco, are commodities which are no where necessaries of life, which have become objects of almost universal consumption, and which are therefore extremely proper subjects of taxation." (Smith, 1776). With few exceptions, tobacco product taxes have been relatively easy to administer and collect, have provided limited opportunities for tax avoidance and evasion, and have generated significant revenues (Sunley *et al.*, 2000; Yurekli, 2002).

In recent decades, as evidence on the health consequences of tobacco use has accumulated, additional motives for tobacco taxation have emerged. Of particular importance is the use of tobacco taxation as a tool for improving public health. This motive has gained prominence as economic evidence emerged on the effectiveness of increased tobacco product taxes and prices in reducing tobacco use, particularly among children and less educated, lower-income populations (Chaloupka *et al.,* 2000a).

Not that long ago, the conventional wisdom was that the addictive nature of tobacco use implied that increases in prices would have little or no effect on use. However, considerable economic research over the past three decades has clearly demonstrated that increases in tobacco taxes and prices are effective in reducing tobacco use. Well over one hundred studies from high-income countries consistently find that a 10% increase in cigarette prices will lead to relatively immediate reductions in overall tobacco use of between 2.5% and 5% (Chaloupka *et al.*, 2000a). About half of the impact on aggregate consumption results from reductions in the prevalence of smoking and half from reductions in cigarette consumption among continuing smokers (Chaloupka *et al.*, 2000a). Growing evidence from low- and middle-income countries suggests that the same price increase reduces overall smoking by up to twice as much (Jha & Chaloupka, 1999; Ross & Chaloupka, 2006). Given the addictive nature of tobacco use, the impact of a permanent price increase will take several years to fully appear, as addicted users respond to the increase in price. Estimates from the USA suggest that the long-run reductions in use resulting from a permanent price increase are about double the short-run effects (Chaloupka *et al.*, 2000b).

The reductions in prevalence caused by tax and price increases are largely the result of increased cessation among current tobacco users. Higher taxes and prices lead

numerous users to try to quit; while many eventually relapse, a significant number are successful in the long-term (Tauras & Chalouka, 2001; Tauras, 2004). In addition, key populations, such as youth and those on low incomes, are particularly sensitive to price. Growing evidence indicates that higher taxes and prices are particularly effective in reducing the number of youth who initiate regular smoking (Tauras *et al.,* 2001; Chaloupka, *in press*). Similarly, as implied by economic theory, tax and price increases lead to greater reductions in tobacco use among low-income, less educated populations than among higher-income, more educated persons (Townsend *et al*. 1994; Farrelly *et al*., 2001). Given current smoking trends, tax and price levels, and evidence on the effects of price on smoking by different age and income groups, estimates indicate that tens of millions of premature deaths, that would have otherwise been caused by tobacco use over the next 50 years globally, could be averted by relatively modest increases in tobacco product prices (Jha *et al*., 2006).

A final, related motive for tobacco taxation is that the tax can be used to correct for the external costs resulting from tobacco use. These include the healthcare costs from treating diseases among nonsmokers, as well as their lost productivity, that are caused by exposure to tobacco smoke, along with the publicly financed healthcare costs to treat tobacco-attributable diseases among tobacco users.

While there has been extensive research on the impact of tobacco taxation on tobacco use behaviours, country-specific evidence is lacking in most countries. In many countries where evidence is available on aggregate relationships, little is known about the impact of taxes and prices on tobacco use among key subpopulations (e.g. youth, low-income persons). Even in countries where substantial research has been done on these issues, questions remain (e.g. on non-linearities on the impact of tax and price on tobacco: whether large tax increases have disproportionately larger or smaller effects than smaller tax increases).

As illustrated in Figure 5.3, the effectiveness of tobacco taxation in reducing tobacco use behaviour and concomitant harm, generating revenues, and covering the costs of tobacco use depends on:

- the degree to which increased taxes raise the prices of tobacco products, including the extent to which tobacco product manufacturers, distributors, and retailers pass along the tax increase, and/or engage in price-related marketing efforts that offset at least some of the amount of the tax increase, as well as the extent to which large-scale smuggling of tobacco products emerges/grows in response to the tax increases;
- the behavioural response of tobacco users to the increased taxes and prices, including not just changes in their tobacco use (e.g. cessation attempts, reductions in tobacco product consumption, compensation), but also changes in their purchasing behaviour (including, for example, switching to cheaper brands, using price reducing promotions, and engaging in efforts to avoid the tax increases);
- the use of the revenues generated from the tax increase to support additional tobacco control activities, such as support for and promotion of cessation interventions (see Section 5.7), and mass media and other public education campaigns (see Section 5.6).

Measuring tobacco tax policy

The first step in assessing the impact of tobacco taxation is developing good measures of the structure of tobacco taxes. There are a variety of taxes that can be imposed on tobacco products. Generally, these include the following types (Table 5.1): customs (import/export) duties, excise taxes, sales taxes, and value-added taxes (VAT).

These taxes can be imposed at different levels and the base for one tax may include the other taxes. In the USA, for example, national excise taxes are collected from tobacco product manufacturers, while state and local excise taxes are collected from distributors. Sales taxes are imposed at the retail level by many states and localities, with most including excise taxes in the base for computing the sales tax. Similarly, the base for the VAT

Types of taxes	Definition
Customs duty	A tax on imports and/or exports, typically applied on a wide range of products, but may include additional levies on particular products.
Excise tax	A tax on selected goods produced for sale within a country or imported and sold in that country; can be specific (based on quantity or weight, independent of price) or *ad valorem* (assessed as a percentage of price).
Sales tax	A tax on a broad range of goods and services sold within a country, generally assessed at the point of sale to consumers and as a percentage of the retail price.
Value-added tax (VAT)	A general, indirect tax on consumption that is applied at each stage of production and distribution based on the value added to the product at that stage.

Sources: Yurekli (2002); Sunley *et al.*(2000)

Table 5.1 Types of Taxes Applied to Tobacco Products

used in many countries includes all excise taxes that have been collected, typically from tobacco product manufacturers.

Tobacco product excises are the most important of these, given that the others are typically applied to a wide range of goods and services, including tobacco products, while excises are applied to a few specific products (e.g. alcohol and gasoline). There are two basic types of tobacco excise taxes: specific taxes and *ad valorem* taxes (Table 5.2). Specific excise taxes are based on some measure of quantity, such as per stick taxes on manufactured cigarettes or weight-based taxes on roll-your-own tobacco. *Ad valorem* taxes are based on a measure of value and are typically applied as a percentage of the price (e.g. 50% of the manufacturer's price). When measuring *ad valorem* taxes, it is helpful to include measures of the monetary value of the tax in addition to the percentage rate that is applied. Most countries apply some mix of specific and *ad valorem* taxes to tobacco products. Finally, for purposes of comparing tobacco taxes across countries, it is useful to express these taxes as a percentage of retail price including, when relevant, as a percentage of price for different categories within a product type (e.g. for locally produced and international brands of cigarettes).

Each form of the excise tax has advantages and disadvantages in achieving the goals discussed above (Sunley *et al.,* 2000; Yurekli, 2002). The revenues generated from specific excise taxes tend to be more stable than those generated from *ad valorem* excise taxes, given that revenues from the latter vary more with industry pricing strategies (e.g. industry price cuts are effectively subsidized by the government when *ad valorem* taxes are applied). In the presence of high inflation, however, the inflation-adjusted value of the revenues from specific excises will fall over time, unless the tax is increased regularly, in contrast to the revenues from *ad valorem* taxes (assuming that industry prices are keeping pace with inflation). Specific excise taxes will generally result in a greater variety of products than will *ad valorem* taxes, since the price difference between higher quality and lower quality products will be smaller with specific taxes, creating a greater incentive to produce higher quality products. In general, if the primary motive for tobacco taxation is to reduce tobacco consumption, imposing specific tobacco excise taxes would be

Construct	Tobacco Product Taxes
Measures	Specific and *ad valorem* excise taxes applied to tobacco products.
Sources	Ministry of Finance, others (e.g. International Monetary Fund, WHO)
Validity	"Gold standard"
Variation	Different types of excise taxes and/or different tax rates are likely to be applied to different types of tobacco products; in some countries, sub-national tobacco excises are important to measure.
Comments	Useful to obtain other measures of tobacco tax administration, such as whether or not tax stamps are required, as well as excise taxes in other nearby jurisdictions. Also useful to estimate tax as a percentage of retail price for comparisons across countries and for assessing impact of tax on price in response to tax increases.

Table 5.2 Measures of Tobacco Product Taxes

preferred, particularly when inflation is relatively low (Sunley *et al.*, 2000).

In evaluating the impact of increases in tobacco product taxes on key outcomes, the size and timing of the increase will be important. For example, large tax increases can be implemented all at once or phased in through a series of more incremental increases over time. Existing estimates suggest a relatively linear relationship between the size of a tax increase and its impact on tobacco use behaviours; more research is needed to assess potential non-linearities in this relationship, differences in the effects of one-time large increases versus a series of smaller increases that add up to an equivalent increase over time, and related issues.

Given that excise taxes are typically included in the base for sales taxes and VAT, it is important to understand how these taxes are applied to tobacco products in order to assess the impact of a tobacco tax increase on the prices users pay for tobacco products. Similarly, other aspects of tax administration will be integral to understanding the impact of these taxes on tax avoidance and smuggling, including: whether or not tax stamps are required and, if so, the design of the stamp and how it is applied; at what stage in the manufacturing and distribution process the taxes are collected; regulation and licensing of those involved in the distribution of tobacco products; the treatment of existing stocks of tobacco products when taxes are increased (e.g. whether or not "floor" taxes are applied); and more (Sunley *et al.*, 2000; Yurekli, 2002). In addition, there are other policies that focus on improving tax compliance, such as policies that target direct sales of tobacco products (e.g. Internet, mail, and phone sales), and that limit or ban duty free purchases. Finally, some policies address the ultimate impact of tax increases on retail prices for tobacco products, such as policies that specify minimum prices for these products or that ban price reducing promotions for them.

In monitoring tobacco taxes and prices over time, it will be important to account for the effects of increases in the prices of other goods and services consumed (inflation). Taxes that are infrequently increased, or that increase slowly relative to the prices of other goods and services, will lose their value over time, potentially resulting in decreases in the inflation adjusted value of tobacco product prices (as, for example, occurred in the USA through much of the 1970s and early 1980s (Chaloupka, *in press*)). Declines in the relative (inflation adjusted) prices of tobacco products, all else constant, will lead to increases in the use of these products.

Proximal variables: measuring tobacco product prices

Understanding how tobacco tax increases affect the prices users pay for tobacco products is critical in measuring the effectiveness of tobacco taxation in both reducing tobacco use and in generating revenues; that is, price is the key mediator for tax. Increases in tobacco taxes are expected to result in increases in the prices of tobacco products. The extent to which tax increases are passed on to tobacco users will be moderated by number of factors, including the structure of the tobacco product market, tobacco industry pricing strategies, the costs of producing tobacco products, the potential for tax avoidance and smuggling, and the extent to which tobacco use responds to changes in prices (Chaloupka *et al.,* 2000a). In countries where the tobacco product markets are dominated by one firm and/or where costs of producing rise rapidly with output, it is likely that an increase in tobacco product taxes will result in less than comparable increases in tobacco product prices, particularly when tobacco use is relatively responsive to changes in price. In contrast, in countries where the tobacco product markets are highly competitive and where per unit production costs are independent of output, increases in tobacco taxes are likely to result in comparable increases in the prices of tobacco products. Existing empirical evidence, largely from the USA, indicates that increases in tobacco taxes result in increases in tobacco product prices that will match or exceed the increase in taxes (Chaloupka *et al.*, 2000a).

A variety of approaches have been used to measure retail cigarette and other tobacco product prices at different levels of aggregation. These approaches differ widely in their cost and coverage. Retail price data can be collected from individuals, households, and retail outlets, and can be aggregated to the market, sub-national (e.g. state or province), or national levels. Some price data may be available from government sources, while others will be available from commercial or other private sources. Costs of obtaining or developing alternative price databases will vary considerably based on source and/or level of detail. Different types of price data are needed to answer different questions. For example, a composite measure of prices is sufficient for analyses that look at the impact of price on aggregate consumption, while brand specific prices will be important for analyzing the effect of relative prices on brand choice. As noted above for tax, it is important to account for the effects of inflation when evaluating the impact of tobacco taxes on tobacco product prices, and of taxes/prices on tobacco use and related outcomes.

For purposes of comparison, alternative retail price collection strategies will be grouped into three categories, based on the form of data collection: technology-based, observational, and survey (Table 5.3). In places where multiple methods have been used to measure price, the measures produced are generally highly correlated with one another and follow consistent trends.

Technology-based systems for measuring prices:

Some measures of prices based on technology-based data collection systems take advantage of sophisticated technologies employed by a growing number of tobacco product retailers in at least some countries. Most prevalent are the “scanner-based” data collection systems that utilise the universal product codes (UPCs) included on most product packaging. These systems are most widely used in high-income countries, but are spreading to many low- and middle-income countries. Other technologies that go beyond those based on UPCs, such as radio frequency identification (RFID) tags, are starting to emerge, but have not yet been widely implemented. Companies such as A.C. Nielsen (http://www.acnielsen.com) and Information Resources International (IRI) (http://www.infores. com) collect and sell these data in a growing number of countries.

These high-tech data collection systems have the advantage of collecting more comprehensive and more detailed data than can be collected using other approaches. They essentially provide a census of the prices paid for every sale, by UPC, in the outlets that employ the relevant

Method	Description
Technology-Based	Uses of Universal Product Code (UPC) and scanner technology (or others) to collect detailed information on the sale of every tobacco product, including information on price, quantity, and use of promotion at detailed product/brand-level; limited to sample of participating vendors with relevant technology. Also used at the household level to collect detailed information on all household purchases of tobacco products and other consumer goods.
Observation	Use of trained observers to collect information (price, price promotions, packaging information, etc.) on selected tobacco products from a sample of tobacco product vendors.
Survey	Use of mail or telephone questionnaires of tobacco product vendors to collect information on prices and price promotions for selected products, or surveys of tobacco product users to collect information on prices and use of promotions for the products respondent consumes.

Table 5.3 Methods for Collecting Tobacco Product Prices

technology. Brand and package-specific information can be extracted from these data, as well as information on a variety of price-related promotions at the retail level. For example, prices for single pack, carton, and any multi-pack specials will appear separately for every brand in these data; to the extent that there are other in-store promotions, such as on-package coupons or other retail value added promotions (e.g. a free gift with cigarette purchase), these will be separately available as well. This type of data was used, for example, to document the associations between retail promotions for cigarettes and the Master Settlement Agreement, state cigarette excise taxes, and state tobacco control programmes in the USA (Loomis, *et al.*, 2006). In addition to the price data, these systems produce good measures of market share and the share of sales that reflect at least some tobacco company promotional efforts in the sample of participating tobacco product outlets. These data were also used to examine how prices of and promotions for premium, discount, and deep discount cigarettes in the USA affected the share of the cigarette market accounted for by each category (Tauras *et al.,* 2006).

Comparable systems use UPCs and in-home scanners to collect data on prices and purchases at the household level from nationally representative samples. In the USA, for example, A.C. Nielsen maintains its HomeScan sample; IRI's comparable sample is the Combined Outlet Consumer Panel. Both are panels of tens of thousands of households that include information on the outlets from which household members purchase various products and the quantities that are purchased; prices are input for purchases from outlets that do not participate in the store level, scanner-based database. Both companies maintain similar databases in other countries, as does Sofres, Taylor and Nelson, Inc. (http://www.tns-global.com).

The major limitation of these systems is their coverage. Given the manner in which data are collected, stores that do not employ the relevant technologies will be excluded. While these technologies are relatively widely used in high-income countries, there are many retailers that do not yet employ them; most likely in many low- and middle-income countries. In addition, at least some large retailers in some countries (e.g. Wal-Mart in the USA) do not participate in the systems. To the extent that prices, promotional activities, and sales patterns differ among included

and excluded outlets, the data produced by these systems may not be representative. The home-based data collections partially fill this gap, but generally do not include representative samples of households at sub-national levels. In addition, these systems do not provide complete geographic coverage, but instead tend to focus their data collection efforts on larger metropolitan areas. Again, to the extent that there are differences in prices, promotional efforts, and sales between more urban and more rural markets, the data produced by these systems may not be representative. In addition, these data are relatively expensive, particularly as the desired information is more disaggregated. Finally, given that these data are provided by commercial vendors, there will likely be some constraints imposed on how the data can be shared and/or published.

Observational approaches to measuring prices:

A second approach to collecting tobacco product price data is the use of observational data collection methods. This approach involves trained observers visiting tobacco product vendors and collecting information on the prices of various tobacco products, as well as measures of promotions that affect the price that consumers pay for these products (e.g. on pack coupons, multi-pack promotions). This approach is generally employed in collecting the tobacco product price data that are included in consumer price indices in many countries. Similarly, the Economist Intelligence Unit (EIU; http://www. eiu.com) uses this approach to collect tobacco product prices (cigarettes and pipe tobacco) in 129 cities around the world. In the USA, ACCRA (formerly the American Chamber of Commerce Researchers' Association) used to collect cigarette prices for 250-300 metropolitan areas each quarter (http://www.coli.org). In addition, some market research companies (e.g. A.C. Nielsen) conduct store observations that collect detailed data on pricing, product placement, in-store advertising and promotion, and other marketing activities.

In these systems, cigarette and other tobacco product prices are typically one component of a larger price data collection effort. The EIU, for example, collects prices on over 160 products. These systems have usually been developed to measure changes in the cost-of-living over time and/or to compare the cost-of-living across locations. The EIU data were used, for example, to compare the affordability of cigarettes among low-, middle-, and high-income countries, and to assess the impact of affordability on cigarette consumption in these countries (Blecher & van Walbeek, 2004). Some of the more proprietary databases are used by companies to track their own pricing and marketing strategies, as well as to obtain information on the strategies employed by their competitors.

Tobacco policy researchers have also employed observational data collection methods to measure cigarette and other tobacco product prices and price-related promotions. For example, the ImpacTeen project employed these methods to collect price and other data from almost 17,500 retail outlets in nearly 1000 US communities from 1999 through 2003 (http://www.impacteen.org). These data were used, for example, to examine the impact of cigarette prices and point-of-sale cigarette marketing on youth smoking uptake (Slater *et al*., 2007). Similarly, the Rockefeller Foundation's Trading Tobacco for Health Initiative (TTHI) developed and pilot tested methods for collecting these data in several Southeast Asian countries, as well as in selected other countries (http://www.tobaccoevidence.net).

There are a number of challenges to employing these methods to develop good measures of tobacco product prices. Perhaps the most significant is the development of the appropriate sample frame for use in selecting a representative sample of tobacco product retailers. Alternative approaches include using business list data (available at some cost from commercial vendors) to identify potential tobacco product vendors, sampling geographic areas and thoroughly canvassing them to identify these vendors, or using convenience samples of vendors that are readily identifiable and easily observed. ACCRA, for example, requires that observers

visit a minimum of five stores, but recommends more (particularly when there is substantial variation in price), but provides limited additional guidance (for details, see the ACCRA manual at http://www.coli.org/surveyforms/colimanual.pdf). In contrast, the ImpacTeen project used business list data to develop a sample of all retailers that might sell tobacco products (based on self-reported Standard Industrial Classification (SIC) codes), then conducted a short telephone screening call with each to determine whether or not they did sell tobacco products, and drew their sample from those that did sell. The TTHI, in contrast, employed a grid search method to canvass given geographic locations to identify tobacco product vendors. To the extent that there are a large number of more informal tobacco product vendors (e.g. street vendors, kiosks, etc. that might not appear in commercial business lists), the latter approach seems most appropriate.

A second challenge relates to the geographic area to be covered by the observational data collection methods. Producing nationally representative price measures in large countries would require multiple teams of observers throughout the country and would involve considerable expense. Alternatively, the approaches used by the EIU and ACCRA that limit data collection to cities or metropolitan areas and that employ convenience samples, will be significantly less costly. However, to the extent that there are significant geographic differences in pricing and price-related promotional efforts, these differences will not be reflected in measures based on data from a limited number of locations.

A third challenge is determining the set of tobacco products for which price and other price-related data will be collected. In contrast to the high-tech methods described above that produce very detailed data at the UPC level, it is not feasible to try and collect data for more than a small fraction of available products. The EIU, for example, collects data on three products: one pack of Marlboro (or another international brand if not available), one pack of a popular local brand, and 50 grams of MacBaren pipe tobacco; similarly, ACCRA used to collect prices for a single product: a carton of Winston king-sized cigarettes. Research-based observational data collection efforts have typically selected a subset of products that includes the most widely consumed products/brands. When there are different price or other categories for some products (e.g. premium and discount cigarettes, or international and domestic cigarettes), then popular products/brands within each category are collected. To the extent that there is limited variation within a given product category (e.g. premium brand cigarettes), measures of price based on observational data collection for a small number of products will be a good reflection of overall prices.

A fourth challenge to developing good measures of tobacco product prices, based on the observational data collection methods, relates to the aggregation of the brand specific data from multiple outlets into a composite price measure. Ideally, this measure would be an average price measure weighted so as to reflect the shares of sales of the different brands that it includes, as well as the sales in different types of outlets (to the extent that there are differences in prices across outlets). Brand share data may be available nationally, but are less likely to be available locally. Similarly, data on the share of sales accounted for by sales in different types of outlets are unlikely to be readily available in many countries.

Survey approaches for measuring prices:

A third approach to collecting data on tobacco product prices and price-related promotions is the use of survey methods. These include mail and telephone surveys of tobacco product vendors and population surveys (including surveys of tobacco users only).

The cigarette price data that have been most widely used in economic studies of the impact of cigarette taxes and prices on smoking behaviour are the price data reported for the USA in the *Tax Burden on Tobacco* (TBOT) (Orzechowski & Walker, 2007). Annual, state level average cigarette prices have been collected and reported for over five decades in the TBOT, with reported prices reflecting weighted averages of prices for single packs, cartons, and vending ma-

chine sales (where weights are based on national shares); since the growth of discount brands in the late 1980s, alternative price series, one including discount brands and one excluding these brands, have been produced. Researchers have used these price data to examine the impact of prices on tax paid cigarette sales (Farrelly *et al.,* 2003a), adult smoking prevalence (Farrelly *et al.*, 2001), smoking cessation (Tauras & Chaloupka, 2001; Tauras, 2004), and youth smoking initiation (Tauras *et al.,* 2001). Reported prices are supposed to reflect the normal retail prices, exclusive of any price-related promotions. The price data are collected through a mail survey of cigarette retailers across the USA. Limited information about the survey itself, sampling frame, response rates, and underlying data is available from internal Tobacco Institute documents (Tobacco Institute, 1991).

In exploratory work on data collection methods done as part of the ImpacTeen project, researchers also conducted a mail and telephone survey of representative samples of tobacco product retailers in three US states, along with observational data collection in representative subsamples in each state. Prices were collected for ten brands of cigarettes in three price categories (premium, discount, and deep discount), as well as for a few other widely consumed tobacco products. In addition to price data, information on various price-related promotions was also collected. As was expected, response rates to the mail survey were very low (less than 10%); response rates to the telephone survey were also low, albeit higher than to the mail survey. However, despite the relatively low response rates, the measures of price produced from the three methods were generally consistent with one another; though there was somewhat greater variance in the measures of the extent of promotional activity.

Similar efforts have been undertaken in other countries. For example, data were used on cigarette prices collected from a commune level survey to estimate the impact of price on the initiation and cessation of tobacco use in Vietnam (Laxminarayan & Deolalikar, 2004). Likewise, cigarette price data were collected from market level surveys in China and Russia to estimate the impact of price on smoking in these countries (Lance *et al.*, 2004).

The use of telephone or mail surveys of tobacco product vendors to collect data on tobacco product prices and price-related promotions faces several of the same challenges as described above for systematic observational data collection. Of particular note are the difficulties in developing an adequate sampling frame (particularly in countries/-markets where more informal vendors are important), the feasibility of collecting detailed data for many products, and the challenges in aggregating the data in order to produce representative price measures.

Alternatively, price and price-related promotions data can be collected through population surveys. A number of cross-sectional and longitudinal surveys have collected information on cigarette prices from respondents. These include population surveys, such as the Global Youth Tobacco Survey (GYTS), which has included questions on price in many of the countries in which the survey has been implemented, and the planned Global Adult Tobacco Survey (GATS) (these surveys are described in Section 4.3). Similarly, the International Tobacco Control Policy Evaluation Study's (ITC) longitudinal surveys of adult smokers, that are being conducted in a growing number of countries, asks smokers how much they pay for cigarettes. Most surveys that inquire about price only ask the relevant questions of current users; some, however, have asked all respondents, while others have asked current and former users.

The price data collected from these surveys are useful in developing aggregate measures of price (e.g. at the national and/or sub-national level, depending on the nature of the sample). However, the use of the individual's self-reported price in analyses that look at the impact of price on respondents' smoking behaviour is problematic given the likely reverse causality between smoking behaviour and price. That is, heavier smokers, all else the same, are more likely to choose less expensive brands, purchase in greater quantities, seek out less costly vendors, engage in tax avoidance, and take

advantage of price-reducing promotions. Given this, treating the self-reported price as an exogenous determinant of individuals' smoking behaviour will lead to an overestimate of the effects of price. Appropriately aggregated measures of price based on individual level self-reported prices can be used to overcome this problem.

In addition to using the surveys to collect prices, it is important to also collect information on the brand that the individual purchased including information on various characteristics of the product (e.g. for cigarettes, length, filter or no filter, and others), and the quantity purchased (e.g. number of cigarettes, grams of smokeless tobacco, etc.); these measures are discussed in detail in Section 3.1. Some surveys use questions that rely on respondents' ability to perform mathematical computations (e.g. on average, how much did you pay for each pack of cigarettes you bought last time?). For respondents that buy by the pack, this is straightforward; it is somewhat more difficult for those who buy by the carton and even more difficult for those who take advantage of multi-pack specials (e.g. buy-three-get-two-free). Alternatively, one could ask how much the respondent paid for their purchase and what quantity was purchased (e.g. for cigarettes, in packs, cartons, single cigarettes, other combinations). For example, the first draft of the GATS questionnaire includes the following questions:

The last time you bought cigarettes for yourself, how many cigarettes did you buy?

INTERVIEWER: RECORD NUMBER AND UNIT BELOW

1. Cigarettes
2. Packs

→ How many cigarettes were in each pack ? _____

3. Cartons

→ How many cigarettes were in each carton? _____

4. Other: Specify: _____

How many cigarettes were in each [FILL]? ___

How much money did you pay for this purchase?

_____ [FILL COUNTRY CURRENCY]

Ideally, the price questions would be asked so as to capture the use of any additional price-reducing promotions (e.g. coupons) at this purchase; the collection of data on use of promotions is described in more detail in Section 5.4. One example of these types of price questions, from the US Current Population Survey's Tobacco Use Supplement (where the majority of purchases are by the pack or carton), is:

What price did you pay for the LAST pack of cigarettes you bought? Please report the cost after using discounts or coupons.

$___.___ ___

What price did you pay for the LAST carton of cigarettes you bought? Please report the cost after using discounts or coupons.

$__ __.__ __

When asking questions about price and purchase-related information, some surveys focus on the most recent purchase (as in the examples above), so as to minimize recall error and get a consistent measure of current prices. Other surveys focus on the "usual" price paid, brand consumed, and other purchase-related information. An example of this is the series of price questions from the US Adult Tobacco Survey:

How much do you usually pay for a pack of cigarettes?

$___.___ ___

How much do you usually pay for a carton of cigarettes?

$___ ___.___ ___

This approach has the advantage of capturing con-sumers' typical behaviour, but will not pick up any changes in behaviour that may be particularly relevant for measuring price (e.g. a smoker taking advantage of a buy-one-get-one-free promotion for a brand other than the usual brand on their last purchase). Some ask questions on both usual and most recent purchase (e.g. some versions of the ITC surveys include variants of both types of questions).

In addition to, or as a substitute for, asking respondents for some of the detailed information on the products they consume, some

surveys have asked respondents to report the UPC on the pack of cigarettes that they are currently consuming (which can be used to determine brand, filter, flavor, length, etc.); the same could be done for other manufactured tobacco products. Likewise, in some face-to-face surveys (e.g. the version of the ITC survey conducted recently in Poland), respondents are asked to show the interviewer the pack that they are currently consuming; the trained interviewers can then record this information, along with other relevant information that can be helpful in assessing the extent of tax avoidance and smuggling (as discussed below). In many countries, price is recorded on the pack; to the extent that this is the case, respondents (or the interviewer) asked to examine the pack can report the listed price.

Some efforts to measure tobacco product prices rely on consumer or household expenditure surveys. These surveys typically collect information on expenditures on a wide variety of goods and services, including tobacco products, consumed by the individual/household over some specified period of time (e.g. previous week, previous month). Some of these surveys also include questions on tobacco product consumption and, in household surveys, who in the household consumed these products. Responses to these questions can be used to estimate price (by dividing total expenditures on tobacco products by total consumption of these products). This type of derived measure of price should be used with more caution than the more direct measures described above given the potential compounding of errors across the various questions. This is of particular concern in household expenditure surveys where one family member reports on overall household expenditures and consumption, and/or in surveys where broad measures of tobacco expenditures and use are reported, rather than measure of product-specific expenditures and consumption.

Researchers have used either self-reported prices or price measures based on self-reported expenditures in a variety of studies. For example, one analysis of the demand for cigarettes in Bulgaria used self-reported cigarette prices (Sayginsoy *et al*., 2002), while another used a measure of price derived from self-reported expenditures to estimate the demand for tobacco in Myanmar (Kyaing *et al.,* 2005).

As discussed above, the ability to use these data to assess how changes in tobacco product taxation affect the price consumers pay for these products will depend on the collection of other key variables. Example questions addressing other issues relevant to price are contained in other sections of this Handbook (e.g. brand choice in Section 3.1, use of promotions in Section 5.4).

Finally, some surveys collect a variety of other information related to tobacco taxation and tobacco product prices. For tobacco tax increases to have a meaningful impact on tobacco use behaviours (e.g. promote efforts to quit or prevent youth from starting to consume regularly), the price increases need to be noticed and of sufficient magnitude to raise concerns in the user. How large the increase needs to be for this to happen, however, is moderated by the user's (or potential user's) characteristics, including their tobacco use. For example, economic theory predicts that low-income persons will generally be more responsive to changes in prices of the goods and services they consume than will high-income persons, given that consumption of each accounts for a greater share of the individual's budget. Empirical evidence confirms that this is the case for tobacco products (Townsend *et al.* 1994; Farrelly *et al.,* 2001).

Developing good measures of this awareness and concern is more challenging than measuring observable variables like tax and price. Nevertheless, a number of population surveys have attempted to address this by collecting data on the role of tax and price changes in an individual's smoking decisions, concerns about tax and price increases, perceptions of responses to increases in prices, responses to specific recent tax/price changes, perceptions about the effectiveness of price increases in reducing smoking (particularly among youth), support for tobacco tax increases, and other related attitudes and beliefs. Little research exists on the relationships of tax and price increases to

these questions and there is little evidence on their validity. A few examples of these types of questions include:

- In the last 6 months, have you spent money on cigarettes that you knew would be better spent on household essentials like food? (ITC)
- In the last month, how often, if at all, did you think about the cost of smoking? (ITC)
- If the price of cigarette rose today by $__.___ per pack, how many cigarettes do you think you would smoke per week? (with comparable questions about switching to a cheaper brand, trying to quit, buying by the carton instead of the pack, etc.) (variations in ITC)
- Did the price of cigarettes affect your decision to stop smoking? (with comparable questions about starting, daily versus occasional smoking, and quantity smoked) (Ontario Tobacco Research Unit Canadian tobacco survey database (OTRU))
- Now thinking about your own patterns of smoking, how much effect on your smoking do you think each of the following would have in reducing your smoking...(a) if the price of cigarettes doubled, would this have a... (OTRU)
- The price of cigarettes has a big influence on keeping people your age from smoking (agreement/disagreement scale) (1999 Florida Anti-Tobacco Advertising/Media Evaluation - State Survey (US-FATMESS))
- Have you talked with friends about the rising price of cigarettes? (US-FATMESS)
- Do you like raising the price of cigarettes to keep people from smoking? (US-FATMESS)
- How much additional tax on a pack of cigarettes would you be willing to support if some or all the money raised was used to support tobacco control programmes? (US Adult Tobacco Survey (US ATS))

These questions can provide data that may be useful for other purposes, but are not of primary importance for evaluating the impact of tobacco taxation (except, perhaps, in some limited circumstances). Questions about support for tobacco tax increases can be helpful in demonstrating public support for these increases, and those that tie support to funding of tobacco prevention/cessation programmes can similarly demonstrate support for these programmes; there are risks to furthering tobacco control, however, if responses indicate a lack of support. Questions on expected responses to tax and price increases can be used to estimate the potential revenue and public health impact of proposed tax increases; these types of questions are common in market research studies, but their predictive validity for tobacco research has not been assessed. Questions about responses to recent tax increases (or decreases, as was the case in Canada in the mid-1990s) can be useful in assessing the impact of these changes, particularly in the absence of comparable baseline data or when attempting to disentangle the effects of tax changes from other policy changes around the same time.

Summary:

Three alternative methods can be used to measure tobacco product prices for use in assessing the impact of tobacco taxation on price and, ultimately, on tobacco behaviours. These methods have different strengths and weaknesses and the cost of implementing each can vary considerably. To the extent that a national measure of price is of most interest and a regularly repeated population survey of tobacco use is in place, including questions on price in such a survey would be the most efficient approach to collecting this measure. Table 5.4 briefly summarizes each.

Distal variables: measuring tobacco product purchase behaviour

To some extent, the impact of tobacco taxation on tobacco use behaviour will depend on opportunities for tobacco users (and potential users) to minimize the effects of the tax increase on the prices they pay for tobacco products. These opportunities will vary from location to location and will depend on factors such as:

- the variety of tobacco products available and the relative prices of these products, given

Construct	Prices of Tobacco Products
Measure 1	Technology-based systems for measuring prices, e.g. "scanner-based" retail sales data, radio frequency identification tags, in-home scanners.
Sources	A. C. Nielsen (http://www.acnielsen.com), Information Resources International (IRI; http://www.infores.com), Sofres, Taylor and Nelson, Inc. (http://www.tns-global.com)
Validity	Clearly validated
Variation	Comprehensiveness of sample varies over time within countries, and will vary considerably across countries as technologies diffuse. Validity will depend on the comprehensiveness of the system.
Comments	More comprehensive data than other approaches (e.g. brand and package-specific information, census of prices paid for every sale, price-related promotion). Limitations include incomplete participation of tobacco product vendors (particularly where there is a large informal sector), limited use of technology in many low- and middle-income countries, incomplete geographic coverage, and relatively high cost of the data.
Measure 2	Observational approaches, e.g. trained observers visit tobacco product vendors and collect price information.
Sources	Economist Intelligence Unit (EIU; http://www.eiu.com), ACCRA (http://www.coli.org), research-based efforts (e.g. ImpacTeen – http://www.impacteen.org), consumer price index, tobaccoproducts component
Validity	Clearly validated
Variation	Existing international systems (EIU) provide limited product, outlet, and geographic coverage. More comprehensive systems could be developed at the country level for expanded set of products, more systematic sampling of vendors, and more representative geographic coverage. Validity will depend on the extent of implementation (e.g. products included, sample of vendors, and geographic coverage).
Comments	There are challenges in getting a comprehensive sample within and among geographic regions. It is also a challenge to determine which prices to assess and how to aggregate across brands. Costs of implementing a comprehensive system are likely to be high in most countries.

Table 5.4 Measure of Tobacco Product Prices

Measure 3	Survey approaches: mail and telephone surveys; population surveys.
Sources	U.S. Tax Burden on Tobacco (TBOT), Global Youth Tobacco Survey (GYTS), Global Adult Tobacco Survey (GATS), The ITC Project
Validity	Clearly validated
Variation	For vendor surveys: inclusion or exclusion of price-related promotions in prices; mix of products on which price/promotion data are collected; sample of vendors included; mail versus telephone survey. Validity will depend on comprehensiveness of survey, sample of vendors, and response rates. For population surveys: focus on last purchase versus usual purchase, quantity purchased, inclusion of price promotions. Validity will depend on the quality of the price questions.
Comments	For vendor surveys: response rates are low; difficult to develop an adequate sampling frame and collect detailed data on many products. For population surveys: self-reported price at the individual level should not be used to study the impact of price on individual level tobacco use behaviours; subnational aggregation of price can be problematic.

ACCRA: Formerly, American Chamber of Commerce Researchers' Association

Table 5.4 Measure of Tobacco Product Prices

the opportunities for substitution from one type of tobacco product to another in response to changes in relative prices that result from changes in taxes (e.g. switching to roll-your-own tobacco in response to an increase in taxes/prices on manufactured cigarettes)

- the variety of brands for a given type of product, particularly brands in different price categories, that allow for switching to less expensive brands in response to increases in taxes and prices (e.g. difference in prices among premium, discount, and deep discount brands; differences in prices between international brands and locally produced brands)
- the availability of "discounts" based on the quantity purchased (e.g. prices for cigarettes that are lower per pack/per stick when purchased by the carton rather than by the pack)
- the availability and extent of industry promotions that reduce the price or provide added value for at least some purchases including: on-pack money off coupons; multi-pack promotions (a different form of quantity discount, such as buy-one-get-one-free promotions); special price reductions at the point of sale; distribution of free cigarettes at sponsored and other events; and value added promotions, such as gifts with purchases (e.g. a "free" cigarette lighter with the purchase of a pack of cigarettes). Some of these will be available at the point of sale, while others may come through other channels (e.g. coupons in print advertising and direct mail promotions) (see Section 5.4)
- differences in prices among local tobacco vendors (e.g. differences in prices between "convenience" stores where a premium is paid for the "convenience") and less convenient, bulk purchase stores where quantity discounts are extensive
- the extent of an "informal" market in tobacco products (e.g. street vendors with no fixed location), particularly as it allows for distribution of smuggled and/or counterfeit tobacco products
- access to lower tax/price jurisdictions and/or distribution channels (e.g. other countries,

tax-exempt jurisdictions, such as Native American reservations in the USA, the Internet, and other direct tobacco product vendors), and ready access to these jurisdictions/channels that allow relatively easy, low cost opportunities to purchase from/through them.

As described in some detail above for measuring price, there are multiple methods for collection of and/or multiple sources for these data. The technology-based systems can provide comprehensive information on the range of products and brands that are sold in different types of outlets and on the relative prices across products/brands, many of the types of industry promotions for them, and/or the quantity discounts that are available on each. However, as discussed above, these databases are limited in several ways, particularly in capturing the full range of tobacco product vendors (most notably those in the informal sector, the Internet, and other direct vendors), and their utility for assessing the tax avoidance that can emerge in response to tax and price increases.

Observational methods can produce similar information on some of these measures. While not providing the extensive detail on product, brand, relative prices, promotions, and sales that is available in the technology-based systems, observational methods can provide at least some measures of the range of tobacco products and brands available and the types of promotions on at least a selected set of these products. On the other hand, they can be applied to many different types of tobacco product vendors (including direct vendors, those in the informal sector, and others that allow for tax avoidance in nearby jurisdictions).

Similarly, information on all of these measures can be collected through surveys of tobacco users. As discussed in Section 3.1, surveys can provide good measures of the types of tobacco products consumed, as well as on brand choice, while the aggregate data described in Section 4.2 can be used to look at the market share for different types of products and/or brands. Section 5.4 describes the use of surveys to measure awareness of and participation in a variety of tobacco industry promotional efforts, including those that impact on the price tobacco users pay for the products they consume.

Purchase quantity:

Buying in greater quantity (e.g. by the carton instead of the pack) can reduce the per unit cost of tobacco products. Many surveys have assessed purchase quantity; some examples of these questions include:

- The last time you bought cigarettes for yourself, did you buy them by the carton, the pack, or as single cigarettes? (ITC)
- Do you usually buy cigarettes by the pack or the carton? (US ATS)

The GATS questions on price above provide a more flexible way of obtaining quantity purchased that can be applied in a wider range of settings than either of these questions. Similar questions can be developed for other tobacco products. As evidenced by these questions, timing of purchase can vary, with questions focusing on most recent purchase, regular/usual purchases, any purchase, and purchases over some specified period (last week, last month); the same will be true for other questions on purchase behaviour (Table 5.5).

Purchase location:

Many recent surveys have included a question or series of questions on purchase location, including type of vendor purchased from and efforts to avoid taxes by purchasing from different jurisdictions (Table 5.6). Given the extensive variation across countries, the response categories for these types of questions will need to be tailored to a given country so as to include responses that capture the full range of vendors and locations available to tobacco users. For example, the following question has been asked of cigarette smokers in recent waves of the ITC survey in Poland (for both last purchase and usual purchase):

Where did you buy your last pack of (or do you usually buy) cigarettes? (Gas station, Hypermarket, Grocery store/deli, Tobacco Shop, News stand/Kiosk, Marketplace (stationary stand/

Construct	Purchase Behaviour - Purchase Quantity
Measure	"The last time you bought cigarettes for yourself, how many cigarettes did you buy?" RECORD NUMBER AND UNIT BELOW 1. Cigarettes 2. Packs → How many cigarettes were in each pack? _____ 3. Cartons → How many cigarettes were in each carton? _____ 4. Other - Specify: _____ → How many cigarettes were in each [FILL]?
Source	GATS (draft questionnaire)
Validity	Evidence of utility, but with limitations.
Variation	Can be developed for other tobacco products; can be simplified where product packaging is standardised; can ask about last purchase or usual purchase quantity. Accuracy of self-report unclear, particularly from questions that limit responses to packs and cartons.
Comments	Important for assessing efforts to minimize price in response to tax increase by buying larger quantities which often reduce the per unit price.

GATS: Global Adult Tobacco Survey

Table 5.5 Measures to Assess Purchase Quantity

fixed seller), Street seller (mobile seller), Over the Internet, Wholesaler, "Black Market," Other)

Versions of the ITC surveys in other countries exclude some of these responses, but include others; for example, the French version asks about purchases outside France, but within the EU, as well as purchases outside the EU. US and Canadian surveys ask about purchases in other states and provinces, respectively, as well as in other countries and on Native American reservations (which are exempt from state/provincial taxes). Most versions of the ITC survey include duty-free shops as an option, and separate convenience stores from other types of stores. The draft GATS questionnaire includes military stores (which are often tax exempt) and vending machines as options, while noting that the list needs to be adjusted to fit the local environment.

Tax avoidance:

Data from the questions on purchase quantity and location, coupled with the price, product, brand, and promotion questions discussed above and elsewhere in this Handbook, are helpful for assessing users' efforts to minimize prices by changing various aspects of their tobacco product purchase behaviour. They are also of some use in measuring the extent of tobacco users tax avoidance (i.e. their efforts to avoid taxes by purchasing their tobacco products in tax exempt locations, such as Native American reservation stores or from direct sales vendors located on reservations, duty free shops, military stores), or from vendors based in lower tax jurisdictions (e.g. in neighboring or nearby countries or sub-national jurisdictions, the Internet, and other direct vendors based in low tax/price jurisdictions) (Table 5.7). Finally, they have some utility in assessing the extent of more organised smuggling (the illegal transportation, distribution, and/or

Construct	Purchase Behaviour - Purchase Location
Measure	"Where did you buy your last pack of cigarettes?" Responses tailored to local environment, can include: gas station, hypermarket, supermarket, grocery store/deli, convenience store, large discount store, tobacco shop, news stand/kiosk, marketplace (stationary stand/fixed seller), street seller (mobile seller), Native American reservation, military store, over the Internet, by mail, by telephone, wholesaler, another jurisdiction (e.g. country, state, province), "black market", others
Sources	The ITC Project, GATS, and other surveys
Validity	Evidence of utility, but with limitations.
Variation	Response categories need to be tailored to the specific country; can ask about last purchase versus usual purchase. Where used, the distribution of responses and trends over time have expected associations with other factors.
Comments	Important for assessing efforts to minimize price in response to tax/price increase by purchasing from lower price vendor.

The ITC Project: The International Tobacco Control Policy Evaluation Study
GATS: Global Adult Tobacco Survey

Table 5.6 Measures to Assess Purchase Location

sale or resale of tobacco products, generally in an effort to avoid all taxes), and/or counterfeiting (production and sale of cigarettes using brand names and packaging of popular brands sold by leading tobacco companies, typically without paying taxes), to the extent that some of the potential vendors will largely be selling smuggled or counterfeit cigarettes (e.g. mobile street vendors selling from backpacks or those in the "black market"). When assessing tax avoidance and smuggling is of particular interest, asking these questions for last purchase, usual purchase, and any purchase over a specified time period (e.g. three or six months), particularly when coupled with information on quantity purchased, can be useful in producing upper and lower bound estimates for the extent of these problems. They can also be useful in assessing the impact of some of the policies designed to increase tax compliance that were mentioned above (e.g. policies targeting the Internet and other direct sales).

Measuring tobacco product smuggling

Given the illegal nature of tobacco product smuggling, measuring its extent for use in assessing the impact of tobacco taxation (both as an outcome and as a factor which may moderate the impact of tobacco tax increases on price and tobacco use behaviour) is more difficult than measuring the constructs described above. While tobacco tax and price levels can help to explain the extent of smuggling, other factors can be as or more important in doing so; these include the degree of corruption in a country and the nature of tobacco product distribution (Jha & Chaloupka, 1999; Merriman *et al.,* 2000; Merriman, 2001). Moreover, improvements in technology, adoption of new policies, and strengthening of enforcement efforts and penalties appear effective in reducing the amount of

Construct	Purchase Behaviour - Tax Avoidance
Measure	Questions on purchase location, quantity, and price described in previous tables
Sources	The ITC Project, GATS, and other surveys
Validity	Evidence of utility, but with limitations.
Variation	Purchase locations relevant for assessing tax avoidance will vary from country to country. Purchasing from other jurisdictions, duty free shops, street sellers, and direct vendors (e.g. on the Internet) will typically reflect efforts to avoid local taxes; some locations will be relevant to assessing smuggling. Where used, associations between these measures and other factors (e.g. local taxes, proximity to lower tax or tax-exempt jurisdictions) are expected.
Comments	Information on differences in prices across vendors can help identify those that may be relevant for tax avoidance, coupled with information on quantity purchased (both last purchase and usual purchase) can provide a range for estimates of the extent of tax avoidance. Will be useful in addressing concerns about loss of tax revenues to tax avoidance in response to tax increases.

The ITC Project: The International Tobacco Control Policy Evaluation Study
GATS: Global Adult Tobacco Survey

Table 5.7 Measures to Assess Tax Avoidance

tobacco product smuggling (Chaloupka *et al.,* 2008). Despite this, concerns about smuggling often emerge as significant barriers to increased tobacco taxation. Developing good estimates of the extent of smuggling can be helpful in addressing these concerns. It is worth noting that counterfeit cigarettes are emerging as a significant component of illicit markets in tobacco products. Some of the methods and measures described in this section will be applicable to assessing the degree of counterfeit as well; for ease of exposition, however, the discussion here will focus on smuggling.

Five alternative approaches to measuring tobacco product smuggling are described in Tool 7 "Understand, Measure, and Combat Tobacco Smuggling" of the World Bank's Economics of Tobacco Toolkit (http://www.worldbank.org/tobacco) (Merriman, 2001). These will be briefly described here (Table 5.8); those interested in applying these approaches should refer to the tool for more details. Some of these have been applied relatively widely, while others have yet to be systematically applied (or even pilot tested).

The first approach is to conduct key informant interviews with relevant industry representatives, law enforcement agents, government officials, and researchers working on these issues to get their estimates of the extent of the tobacco product market accounted for by smuggling. Market research firms have used this approach and published estimates of the share of the market accounted for by smuggled cigarettes (e.g. Market Research International has published these in the *World Tobacco File*). Researchers have linked these data to potential determinants of smuggling (e.g. tax or price levels, corruption (Merriman *et al*., 2000)), and the resulting estimates suggest that the measure produced from the key informant interviews are useful in comparing across countries. When aggregated, estimates of global smuggling produced from these data are consistent with those produced from other methods described below, sug-

Construct	Tobacco Product Smuggling
Measure 1: Key Informant Survey-based Estimates	Surveys of industry representatives, law enforcement agents, government officials, and researchers to obtain their estimates of the extent of tobacco product consumption accounted for by smuggled products.
Measure 2: International Trade Data-based Estimates	Comparison of import and export statistics to determine extent to which exported products do not appear as imports in the countries they were shipped to; utility at the country level is unclear.
Measure 3: Estimates Based on Comparison of Tax Paid Sales Data and Self-report Survey Data	Difference between estimated total consumption from self-report survey data and tax paid sales data can provide estimate of combined tax avoidance and smuggling; accuracy will depend on biases in both and on changes in biases over time.
Measure 4: Estimates Based on Econometric Modeling of Demand for Tobacco products	Use of tax paid sales data to estimate demand for tobacco products, controlling for key Econometric Modeling of determinants (e.g. price, income, policies) and including measures of potential for tax avoidance and smuggling. Accuracy of estimate will depend on quality of data, ability to control for key determinants of demand, and the ability to measure potential determinants of tax avoidance and smuggling.
Measure 5: Estimates from population Surveys	Surveys to identify users' tax avoidance efforts through questions on purchase location and price, can also include efforts to have survey respondents and/or interviewers report on aspects of packaging including tax stamps, warning labels, and other labeling/markings on pack.
Measure 6: Estimates Based on Observation of Tobacco Product Vendors	Observation of tobacco product vendors to look for tax stamps, warning labels, and other labeling/markings on pack in effort to identify smuggled products.
Source	Most methods are described in World Bank's Economics of Tobacco Toolkit: Tool 7 "Understand, Measure and Combat Tobacco Smuggling" (Merriman, 2001).
Validity	Evidence of utility, but with limitations.
Variation	Alternative methods likely to produce different estimates of the extent of tobacco product consumption accounted for by smuggling. Where multiple methods have been used, resulting measures are generally correlated with one another and have the expected associations with other factors (e.g. corruption).
Comments	Most methods have not been applied widely and more research is needed to determine the validity of the estimates they produce. A combination of methods is likely to be needed to obtain good estimates of the extent of consumption accounted for by smuggling. Good estimates will be important in addressing concerns over the extent to which smuggling will emerge/grow in response to tobacco tax increases.

Table 5.8 Measures for Assessing Tobacco Product Smuggling

gesting that they are valid at some level. Whether or not they provide accurate country level estimates has yet to be fully assessed.

A second approach described in the World Bank's smuggling tool, as well as discussed in Section 4.2 of this Handbook, is the use of international trade data to track smuggling. This approach looks at differences between a country's reported tobacco exports to other countries and those countries reported imports. This approach is useful in assessing the extent of smuggling globally, but is of limited utility for gauging the extent of tobacco product consumption accounted for by smuggled products at the country level, given that one can not identify where the products that "disappear" in transit end up being consumed. Some have assumed that they end up in the country that they were destined for based on reported exports, but this is a tenuous assumption at best. At the global level, estimates produced by this approach are comparable to those produced from the key informant approach.

A third approach is the comparison of data on tax paid tobacco product sales and national estimates of tobacco product consumption based on self-reported survey data. To the extent that there are no reporting biases in either, differences between tax paid sales and reported consumption will reflect the combination of organised smuggling and individual tax avoidance. As described in other sections of this Handbook, there may be systematic biases in both the tax paid sales data (Section 4.2) and the survey data (Section 3.1) that can limit the utility of this approach. However, as discussed in the World Bank tool, to the extent that these biases are constant over time, changes in the difference between the two measures can be assumed to reflect changes in tax avoidance and smuggling. However, to the extent that the biases in the two measures change over time and to differing degrees, this approach will be less useful in measuring trends in tax avoidance/ smuggling.

A fourth approach is to use the tax paid sales data to model the demand for tobacco products, controlling for key determinants of sales (e.g. prices, incomes, other tobacco control policies) and including variables that measure the opportunities for tax avoidance and smuggling. These variables would reflect the extent and ease of access to lower tax/price jurisdictions (e.g. extent of Internet access, price differences between neighboring countries, distribution of population near borders, extent of travel between countries), corruption, and other variables associated with tax avoidance and smuggling. Estimates from these models can be used to produce estimates of the extent of tax avoidance and smuggling by predicting what tax paid sales would be if these variables were set to zero. Several studies in the USA, for example, include measures that reflect the differences in taxes or prices between USA states, weighted by state populations and distances from state borders (Farrelly *et al.*, 2003a). Others have looked at this issue across countries (Merriman *et al.*, 2000). The World Bank's smuggling tool provides a detailed step-by-step explanation for using this approach.

The final approach described in the World Bank smuggling tool is to use population surveys to try and identify the extent of use of smuggled tobacco products. The question(s) on location of purchase described above provide some information that can be useful in assessing the extent of consumption accounted for by smuggled products (e.g. based on purchases in the "black market" or purchases from vendors more likely to sell smuggled products, such as mobile street vendors).

Some surveys have gone further in trying to identify consumption of smuggled products. As briefly noted above, this is done by asking survey respondents or, in face-to-face surveys, interviewers to examine the package from which the user is currently consuming for specific features that can indicate whether or not local taxes were paid on the product. Information on the presence or absence of a tax stamp, presence or absence of local warning labels, and other package labeling (e.g. that indicates where the product was intended for sale or that reports tar, nicotine, and carbon monoxide) can be collected. This approach, in part, depends on whether or not tax stamps,

warning labels, and/or other markings are required on tobacco product packaging and on one's ability to link these to specific countries; something that seems reliably done by trained interviewers rather than by survey respondents. For example, in recent waves of the ITC Poland survey, interviewers have been trained to recognize Polish tax stamps, warning labels, and tar/nicotine/carbon monoxide content labels, as well as those from the Ukraine, Belarus, and Russia; if observed stamps/labels are from another country, this is recorded and the country identified, if possible. This approach depends on users' willingness to produce the package from which they are currently consuming and on the respondent's or interviewer's ability to report this information. In the ITC Poland survey, the vast majority of smokers have produced the pack from which they are consuming and interviewers appear to be successfully recording relevant information.

A related approach that is not discussed in the World Bank's smuggling tool, but that has been pilot tested in limited settings, builds on the observational data collection methods discussed above. Observers can be trained to recognize local and foreign tax stamps, warning labels, and other package labels/markings, and can collect this information on packages available for sale in the outlets observed when collecting price, promotion, and other data. This approach has been used to identify smuggled cigarettes in a small convenience sample in Vietnam (Joossens, 2003) and in a pilot study in Poland (http://www.tobacco evidence. net/activities_workshop.html), but has not been systematically applied at the national level in any country.

As the discussion illustrates, each of these approaches has limitations and none will provide "the" definitive measure of smuggling. Each approach needs to be validated and refined; however, together they are likely to produce a good measure of the extent of tobacco product smuggling (Table 5.8).

Incidental effects: fairness of tobacco taxes

The burden of tobacco taxation on the poor (regressivity of the tax) is often raised as a concern in debates over tobacco tax increases. Evaluating the impact of tobacco taxation and increases in tobacco taxes on equity can be helpful in addressing this concern. Equity (or fairness) is a key consideration in the development of any tax policy, including tobacco tax policy. Economists generally consider both "horizontal equity" and "vertical equity" when looking at tax policy. Horizontal equity implies that individuals with the same income should pay the same tax, while vertical equity suggests that those with the greatest ability to pay (those with higher incomes) should pay more in taxes that those with lesser ability to pay. Tobacco taxes in all or nearly all countries are likely to violate the principle of vertical equity, implying that these taxes are regressive (account for a higher proportion of total income for low-income persons). This results, in part, from the greater concentration of tobacco use among less educated, lower-income persons in most countries. Even in countries where tobacco use increases with income, the increase is unlikely to be proportional to income, implying that the share of income accounted for by tobacco taxes falls as income rises. However, several observers have noted that while tobacco taxes may be regressive, tobacco tax increases can be "progressive" given that tobacco use among the poor falls more sharply when taxes and prices are increased than it does among those on higher incomes, so that a greater share of the increase is paid by higher-income consumers (Chaloupka *et al.*, 2000a). Moreover, the equity implications of tobacco taxes should not be considered in isolation, but rather as part of the overall fairness of a country's fiscal system, which will depend on the distributional effects of other taxes as well as of government spending. For example, to the extent that the new revenues generated by tobacco tax increases are used to fund tobacco cessation programmes targeting the poor (e.g. subsidizing treatment and counseling for low-income users) and to support other progressive programmes, concerns about the

burden of the tax increase on the poor are at least somewhat alleviated. This approach was used in the USA for example, where revenues generated from cigarette tax increases have been used to support the expansion of the state Children's Health Insurance Programme, which provides health insurance for low-income children.

Evaluating the equity implications of tobacco taxes and tax increases is typically a complicated exercise. Those interested in assessing the equity implication of tobacco taxation are encouraged to see Tool 6 "Equity Issues, Tobacco, and the Poor" of the World Bank's Economics of Tobacco Toolkit, which provides detailed, step-by-step methods for doing this (Peck, 2002).

Summary and recommendations

This section focused on the measures that are needed for evaluating the impact of tobacco taxation, a highly effective tool for reducing tobacco use. The impact of tobacco taxes on tobacco use behaviours (see Sections 4.2 and 3.1) is mediated by tobacco product prices, tobacco company price-related marketing efforts (see Section 5.4), tobacco users' purchase behaviour, tax avoidance, and smuggling.

Measuring tobacco product taxes is straightforward (see Table 5.2), with information on the level and structure of these taxes readily available from the Ministry of Finance and other sources (e.g. the International Monetary Fund, the WHO's *Global Tobacco Control Report*). In some countries, it will also be important to measure subnational taxes. Three methods for measuring tobacco product prices were discussed in this section: technology-based, observational, and survey-based. These methods have differing strengths and weaknesses and their costs will vary considerably (see Table 5.4). To the extent that a national measure of price is of most interest and a regularly repeated population survey of tobacco use is in place, including questions on price in such a survey would be most efficient. Measuring tobacco product purchase behaviour can be easily done through the addition of a limited set of questions to this survey (see Tables 5.5 and 5.6 for recommended measures). Developing accurate measures of tax avoidance and tobacco product smuggling is more challenging and the validity of these measures is unclear and needs further research. Some of the questions on purchase behaviour in population surveys can be used to provide a range for the extent of tax avoidance (see Table 5.7). Multiple methods, most of which have not been widely applied and which need further research, can be used to assess the extent of tobacco product smuggling (see Table 5.8).

5.2 Measures to assess the effectiveness of smoke-free policies

Introduction

Article 8 of the FCTC, calls for greater protection from exposure to tobacco smoke (Figure 5.4). In the 1980s, some countries began to implement subnational smoke-free policies. By 2004, Ireland, Norway, and New Zealand were the first countries to implement comprehensive smoke-free worksite policies that also included restaurants and bars. Motivated in part by the FCTC mandate to expand smoke-free policies, other countries have followed suit, but the vast majority of nations have not made progress in this area. Understanding if these policies are effective in achieving their goal of reducing exposure to secondhand smoke and improving health outcomes, is important not only for policymakers in places that pass smoke-free policies, but also to help inform policymaking in other jurisdictions.

The main goal of smoke-free policies is to reduce secondhand smoke exposure and thus to improve health outcomes. There are several measures that should be considered when assessing the effectiveness of smoke-free policies, and factors that might influence how the policy may contribute to reductions in secondhand smoke exposure, as well as more distal outcomes related to secondhand smoke beliefs, attitudes, and practices. Furthermore, there are also potential incidental effects of smoke-free regulations, such as possible business losses/gains, and increased cessation activity among smokers.

There is value to assessing constructs around smoke-free initiatives, both before, during, and after their introduction as policy. Before they are introduced in a jurisdiction, the main variables of interest are an inventory of the level of existing smoke-free policies, as well as the belief about the health harms, and attitudes to restrictions in various locations. During the early implementation period of smoke-free policies, variables of interest are those associated with compliance with the policy and how this relates to secondhand smoke (SHS) exposure. During post-policy introduction, these variables remain of interest, but there are others including how health and economic indicators may have or have not changed. Understanding each of these areas is useful for evaluation purposes and helps to guide subsequent policymaking.

Figure 5.5 presents the logic model guiding the constructs discussed in detail in this section. First we need to understand the nature of the policies. What areas are covered and are there exemptions or possible loopholes? Within a jurisdiction, there may be local policies (from local government), or business-specific policies that need to be considered.

The next step is to consider the impact of these policies on markers of exposure to SHS, which is the

> Parties recognize that scientific evidence has unequivocally established that exposure to tobacco smoke causes death, disease and disability. Each Party shall adopt and implement in areas of existing national jurisdiction as determined by national law and actively promote at other jurisdictional levels the adoption and implementation of effective legislative, executive, administrative and/or other measures, providing for protection from exposure to tobacco smoke in indoor workplaces, public transport, indoor public places and, as appropriate, other public places.
>
> WHO (2003)

Figure 5.4 WHO FCTC Article 8: *Protection from exposure to tobacco smoke*

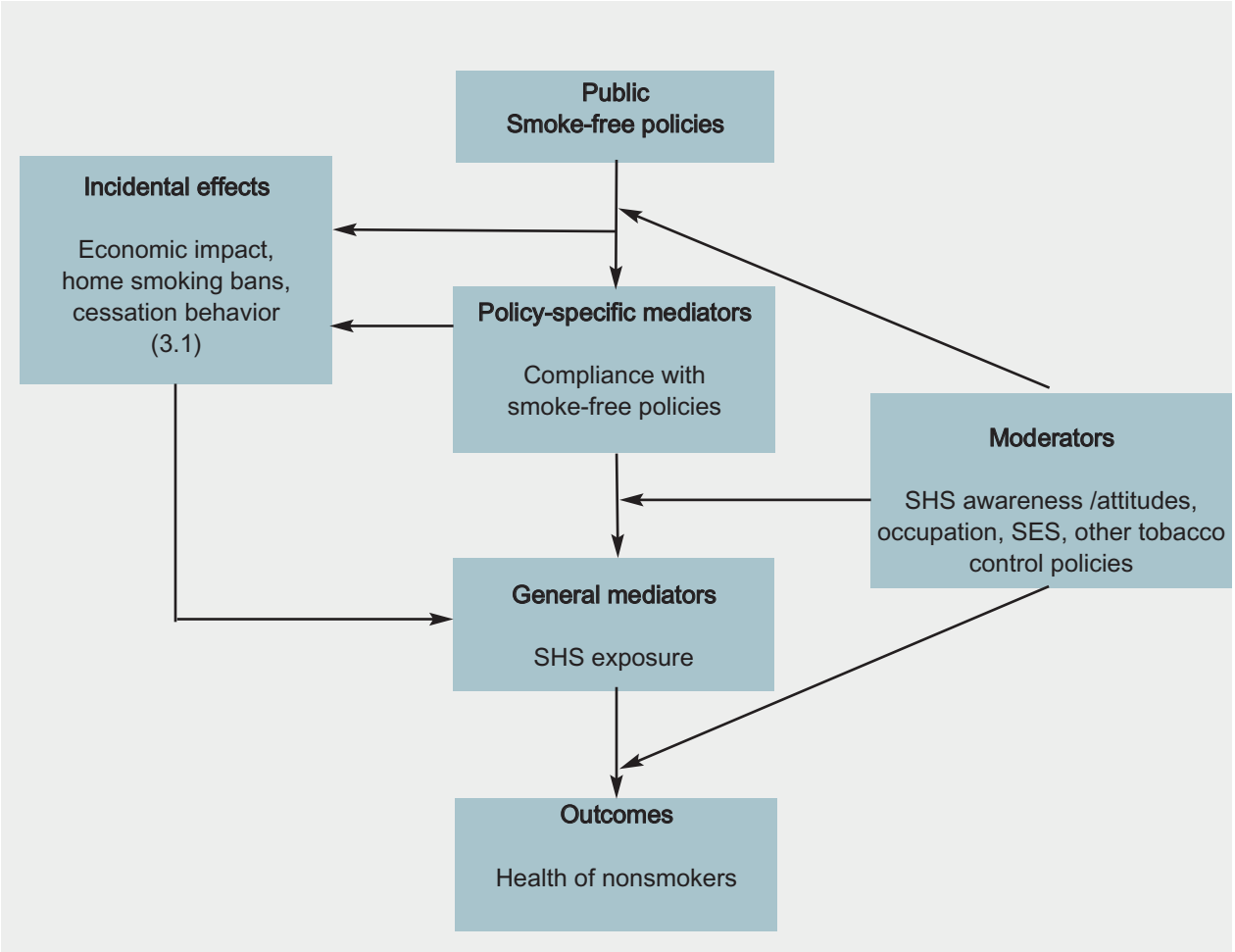


Figure 5.5 Conceptual framework for the evaluation of smoke-free policies

Numbers in parentheses indicate section in the volume covering the topic
SHS = secondhand smoke
SES = socio-economic status

key proximal variable of interest. Compliance with the policy is critical at this point in the model, as poor compliance will weaken the public health benefit of the smoke-free policy, and could even result in a backlash where policymakers overturn the policy because it is ineffectual.

More distal variables that may change in response to smoke-free policy implementation include: people's beliefs about the dangers of SHS, their opinions about the social norms of smoking in different places, as well as the translation of these beliefs into changes in their personal choices regarding rules about smoking in their own personal spaces, such as their home and car. For example, local, grass roots movements in scores of communities in California waged a public information campaign, which led to the passage of local-level clean air policies. Policies can change social norms and beliefs and vice versa.

The primary goal of smoke-free policies is to protect the health of nonsmokers. The greatest benefits should be experienced by those who previously had the greatest exposure. For example, bartenders and wait staff, who previously worked in smoky environments, would derive greater health benefits that a stay-at-home mother or an employee whose worksite had already been smoke-free.

There may also be some incidental effects that need to be rigorously studied in order to address concerns about the impact of these policies. One concern that is raised in nearly every policymaking debate about the merits of smoke-free policies, is that its implementation will adversely impact the economy, as smokers will stop dining out and going to bars. Often this is *the* central issue of the debate and credible information addressing this point needs to be obtained. Some potential economic issues that might be worth considering are the cost savings due to employees' decreased health care costs, increased worker proactivity, and decreased establishment maintenance costs. The other key incidental impact is that smoke-free policies reduce cigarette consumption in smokers. From the public health perspective, this is a beneficial incidental impact, but not the reason why smoke-free policies are considered.

Lastly, there is an array of potential moderating variables to consider for a thorough evaluation. For example, as previously mentioned, one's occupation will moderate the impact of a smoke-free policy. The list of moderator variables presented is not exhaustive, but is meant to provide an overview of additional variables an evaluator should consider. More details on relevant moderating variables are presented in Section 3.2.

Smoke-free policy measures

Through the FCTC mandate, countries are obligated to push for stronger legislation protecting workers and the public from SHS. This is usually accomplished through the passage of policies restricting where smoking can occur in public environments. In some countries, this might mean something as simple as requiring hospitals to provide a smoke-free indoor environment, while others have adopted comprehensive regulations that prohibit smoking in all indoor workplaces, including bars and restaurants. Going beyond the mandate in Article 8 of the FCTC, some jurisdictions are pushing for outdoor smoke-free rules that apply to beaches, entryways to buildings, and parks, for example. In addition to these government mandated policies, individuals or businesses may also adopt voluntary smoke-free policies in their homes and workplaces, irrespective of government policy, although these are not the focus of this section. A summary of commonly used approaches to measure smoke-free policies is given in Table 5.9.

The advantages of assessing policies directly are that their documentation is relatively simple to obtain, and their stipulations provide a standard to be validated against individual exposure data. The negative implications are that the implementation of policies does not always correlate well with actual exposure, due to poor compliance and enforcement. These policies only cover public spaces, and measuring them can get complicated in countries with sub-national policy activity.

Policy-specific mediators or proximal measures – compliance with smoke-free policy

Three types of smoke-free policy compliance measures are summarized in Table 5.10: 1) self-report of policy type implemented; 2) direct observation of com-pliance; and 3) government enforcement and compliance records.

Self-reported measures of exposure can provide a simple measure of the impact of a smoke-free policy. Following implementation of a comprehensive smoke-free policy, the percent of people who report that their workplace is smoke-free should go up and the percent of people who report seeing smoking the last time they went to a restaurant, for example, should go down. These measures are a proxy for the actual smoking policy, as shown in Table 5.9, but are also a key indicator of compliance with the policy, and are presented as such in the model in Figure 5.5. These data are relatively inexpensive to collect if there is an

existing survey in place in the relevant country, state/province, or community, where questions can be added, and the survey can provide for population-based measures of policy impact on compliance. While this measure may lack precision in terms of the extent of compliance, it does provide a useful barometer of the relative compliance levels. We also note that it is important to have pre-policy data, as well as post-policy data, so that the change in compliance can be assessed. For example, post-policy, 20% of people might report that they saw smoking the last time they went to a bar. That might seem high, but if the pre-policy data showed 100% reported seeing smoking in bars, then it demonstrates a dramatic improvement while pointing to areas where programmatic efforts to further increase compliance should be placed. We are not aware of studies that have directly validated these specific self-reported measures with atmospheric measures of SHS or biomarkers of exposure. Observational studies of compliance (i.e. when an independent observer assesses if smoking is occurring in a venue) have been validated (see subsequent sub-section), and the difference in pollution levels is dramatic between smoke-free and smoking-observed venues.

In contrast to self-reported measures of compliance, observational studies *may* provide a more reliable measure of compliance. Field staff are able to observe the presence of evidence of smoking, such as ashtrays or cigarette butts, in such studies. The key element to consider is the design of the observational study. Results may be biased if the venue selection is not random and assessments are made at times that are not representative of typical activity levels. For example, doing an observational compliance study in bars by sending field staff to these locations during weekday afternoons will likely overstate compliance, while performing these checks only during peak times in the late evening will understate compliance. These studies may also not be as generalizeable as self-reported data unless a large, random sample of venues is observed, which can be resource intensive.

Measure	**Smoke-free air policies in key locations**
Sources	Government records; The Americans for Non-smokers Rights Foundation; Smoke-free Lists, Maps, and Data (http://www.no-smoke.org/goingsmokefree.php?id=519 accessed January 25, 2007); CDC State Tobacco Activities Tracking and Evaluation (STATE) System (http://apps.nccd.cdc.gov/statesystem/ accessed January 25, 2007); WHO Global Tobacco Control Report (Shafey *et al.,* 2003)
Validity	"Gold standard" for measuring policy itself, but a strong policy may not translate to low SHS exposure.
Variations	Details of the policies, such as the locations covered, exemption, enforcement authority, and penalties for non-compliance should be tracked unless it proves to be too difficult. National and state/provincial policies are easier to track than local level policies, as there may be thousands of individual sub-national policies to track.
Comments	Tracking national policy will miss local level policy action, as well as voluntary policies passed by businesses and individuals. It may be important to track sub-national policies in some countries.

Table 5.9 Commonly Used Approaches to Measures Smoke-free Policies

Construct	(a) Self-Reported Measures
Measure	Self-reported policy in these areas. Examples of questions include: (Source: ITC Survey) "Which of the following best describes the smoking policy where you work?" (Smoking is not allowed in any indoor area, Smoking is allowed only in some indoor areas, or Smoking is allowed in any indoor areas) (Source: Global ATS) "Which of the following best describes the indoor smoking policy where you work?" (Smoking is not allowed in any indoor areas, Smoking is allowed only in some indoor areas, No rules or restrictions, No indoor areas) (Source: ITC Survey) Public Places – "Which of the following best describes the rules about smoking in drinking establishments, bars, and pubs where you live?" (Smoking is not allowed in any indoor area, Smoking is allowed only in some indoor areas, No rules or restrictions) (Source: Global Adult Tobacco Survey) "During the past 7 days, did anyone smoke in the following indoor places that you visited? " YES NO DID NOT VISIT a. Government buildings or offices? 1 2 3 b. Health care facilities? 1 2 3 c. Schools or universities? 1 2 3 d. Private workplaces? 1 2 3 e. Bars or night clubs? 1 2 3 f. Restaurants? 1 2 3 Example question asked of individuals: (Source: ITC Survey) "The last time [you visited a bar/restaurant/etc.], were people smoking inside the pub or bar?" 01 – YES 02 – NO Example question asked of business owners: (Source: New York City Restaurateur Survey) "Is smoking allowed anywhere in your [restaurant/bar/etc.]?" 1 Yes 2 No
Sources	Questionnaires; for example, Hyland *et al.*, 1999a ; Bauer *et al.*, 2005 ; Borland *et al.*, 2006a ; Borland *et al.*, 2006b ; Fong *et al.*, 2006b
Validity	Evidence of utility. No direct validity study of these self-reported measures, but observational studies assessing the same construct have been validated and show dramatic differences in pollution levels between smoke-free and smoking-observed venues in a variety of settings (see Leaderer *et al.*, 1994; Repace, 2004; Travers *et al.,* 2004).
Variations	Questions can be adapted to specific places of interest. Items reporting the observance of smoking in various places may underestimate exposure if actual smoking not observed.

Table 5.10 Measures of Compliance with Smoke-free Policies (Proximal Variables; Policy-specific Mediators)

Comments	This is a more direct measure of exposure than knowledge that a policy is in place (policy-only data), relatively easy to obtain data, adaptable to address many specific locations, as relevant by each specific policy, but still not a direct measure of actual SHS exposure.
Construct	**(b) Direct Observation of Compliance Measures**
Measure	Observer assessments/spot checks of compliance with smoke-free regulations.
Sources	In person assessments; Hyland *et al.*, 1999a ; Weber *et al.*, 2003 ; Skeer *et al.*, 2004; Engelen *et al.,* 2006
Validity	Clearly valid. Studies have shown large differences in indoor air pollution by type of smoking policy in workplaces, restaurants, bars, and homes (Leaderer *et al.*, 1994; Repace, 2004; Travers *et al.,* 2004).
Variations	The study can be adapted to observe compliance in specific locations as needed by the investigator.
Comments	This is an excellent approach to assess compliance and, budget permitting, should be strongly considered. To obtain a true compliance assessment, ratings need to be done at all hours and on all days, which can increase costs and raise issues of observer safety in some instances. Field work coordination may be more difficult, as observers are often geographically varied in location.
Construct	**(c) Records of Government Statistics on Violations, Enforcement, and Compliance of Smoke-free Policies**
Measure	Government records on the number of complaints, number of enforcement checks, violations, and fines collected.
Sources	Government records; Hyland *et al.*, 1999a; Engelen *et al.,* 2006
Validity	Face validity with concerns noted below.
Variations	Available data may depend on the reporting systems and available data from different jurisdictions.
Comments	High violation rates could be a function of strong enforcement efforts and may not indicate better true compliance. Complaints are also an indirect measure of compliance and the type of complaint needs to be considered (e.g. are complaints from nonsmokers upset about people smoking, or from smokers upset about not being able to smoke inside?).

Table 5.10 Measures of compliance with smoke-free policies

Government enforcement and compliance records are another way to assess observance of smoke-free policies. These are not recommended as the sole source for evaluating compliance, but they can provide useful complementary data when used in conjunction with other exposure assessment data sources. The advantage of these data is that they may be readily available and easy to use. Information typically obtained includes the number of complaints, enforcement operations, and amount of fines collected. Caution must be maintained, as high levels of complaints and violations do not necessarily indicate that the policy is not working well, and in fact, just the opposite may be true. Jurisdictions that take an active role in dedicated enforcement of

smoke-free policies will find more violations, and often the real threat of punishment if caught violating the law encourages greater compliance in the future. It is also important to consider the nature of the complaint. Complaints from those who are upset at smoking occurring where it is forbidden by the policy are much different than complaints by those who are upset with the policy itself.

General mediators or intermediate measures: secondhand smoke exposure measures

Two commonly used sources of actual SHS exposure measures have been previously reported; atmospheric studies, including airborne particulate concentration and nicotine studies, as well as biomarkers studies of exposure (see Table 5.11). Studies testing for the presence of nicotine in the air have the advantage of being specific to tobacco smoke exposure, but nicotine is not assayed in real-time and estimates will only tell about average exposure over time. Particulate matter concentration studies are not specific to tobacco smoke, as other factors like pollution and cooking fumes emit particles, but the presence of tobacco smoke is the dominant source of particulate matter in most cultures even in the presence of high levels of background pollution found in some parts of the world. In these studies the data collection methods allow for real-time particulate concentration data collection. Another advantage with particulate concentration data is that measurements can be compared with well established standards for outdoor air, which aids in communicating results to the public. For example, the average level of particulate matter observed in these types of studies conducted in bars is well above the peak reading experienced during the largest forest fire in the USA State of Colorado's history, which was a 24-hour average $PM_{2.5}$ concentration of 200 micrograms/cubic metre. This compelling imagery is powerful when discussing the risks of SHS exposure and the benefits of smoke-free policies. Both measures (airborne particulate and nicotine concentrations) can complement each other and selection of one measure over the other depends on the questions being asked in the evaluation and resources available. Regardless of which approach is considered, these data are often only collected in a small number of locations because of resource issues (i.e. expense and expertise), but such data can round out exposure assessment data obtained from other sources.

Perhaps the scientific "gold standard" for assessing changes in SHS exposure is examining changes in biomarkers of exposure (Hecht, 2004). Two biomarkers used specific to tobacco smoke exposure are cotinine, a by-product of nicotine metabolism, and 4-(*N*-nitroso-methylamino)-1-(3-pyridyl)-1-butanol (NNAL), a potent tobacco specific carcinogen. Cotinine is typically measured in the urine, saliva, hair, or blood, and NNAL is commonly measured in the urine. Levels of these two biomarkers should be zero if unexposed to SHS, while any detectable level indicates SHS exposure. One methodological approach to collecting this type of data is to couple it with a particulate matter monitoring study where urine samples are collected from nonsmoking field staff before and after spending an evening taking measurements in smoky venues. The change in cotinine and NNAL give a measure of exposure after even a short-term visit. The finding of potent tobacco smoke carcinogens in the urine that were absent prior to going into the field provides a powerful communication message. After a smoke-free policy has been implemented, the cotinine and NNAL measurements in field staff (taken at the beginning and at the end of their work shift) would be expected to show little difference, if any. The main disadvantage of this type of study is the high cost and requirements for adequate facilities to handle storing samples; hence, results are not broadly available precluding much needed comparisons. However, if resources are available a biomarker study can provide very compelling evidence of the real impact the smoke-free policy has on SHS exposure.

Primary outcome of interest – health in nonsmokers

The primary health outcome expected to change following the

Construct	(a) Atmospheric Secondhand Smoke Monitoring
Measure	Direct measurement of particulate concentrations and nicotine levels in ambient air. Particle concentrations can be assessed using a light scatter device and nicotine concentrations can be assessed using a small portable badge that is placed on site for a period of time and sent to a laboratory for chemical analysis
Sources	Roswell Park Cancer Institute Tobacco Free Air website (www.tobaccofreeair.org); Hammond, 1999; Navas-Acien *et al.,* 2004; Repace 2004; Travers *et al.*, 2004; Nebot *et al.*, 2005
Validity	Clearly valid. A Norwegian study showed a strong correlation between ambient particulate matter and air nicotine concentrations (r=0.83) (Ellingsen *et al.*, 2006). One study showed cotinine levels decreased in 35 hotel workers by 69% after implementation of a smoke-free law, while air nicotine levels decreased by 83% (Mulcahy *et al.*, 2005).
Variations	Specific venues tested can vary depending on the policy.
Comments	Real-time assessment of particle concentrations is relatively inexpensive if many samples are being examined and can be compared to benchmarks for outdoor air quality; however, it is not specific to SHS. Nicotine monitoring is specific to SHS levels, but may be more costly than particle monitoring if large samples are collected and does not provide real-time data. Results are often very simple and effective in communicating with the public and policymakers.
Construct	**(b) Biomarkers of Exposure**
Measure	Urine, saliva, or blood cotinine levels provide most direct assessment of SHS exposure. NNAL, a tobacco specific carcinogen, can also be examined in the urine.
Sources	Anderson *et al.*, 2003; Mulcahy *et al.,* 2005; Farrelly *et al*, 2005a; Engelen *et al.*, 2006
Validity	Considered the “gold standard” to which other assessments measure up to.
Variations	Can be combined with particle or nicotine monitoring study to provide a more detailed assessment of what is in the air as well as in the body.
Comments	Most direct SHS exposure assessment. Can be difficult and expensive to obtain, does not rule out other sources of nicotine exposure. Helps to demonstrate the need for stronger SHS policies and to evaluate impact of a policy. Particularly effective in communicating to policymakers.

Table 5.11 Secondhand Smoke Exposure Measures (Intermediate Measures; General Mediators)

implementation of smoke-free air policies is improved health in nonsmokers. A variety of approaches have been used to assess this, and we focus here on items that are not previously presented in Section 3.1. Some studies have relied on self-reported respiratory symptoms collected from large population-based samples (Wakefield *et al.,* 2003a; Lam *et al*., 2005; Ho *et al.,* 2007). This has the advantage of providing more representative data; however, self-reported data are not validated, and the health significance of the report of fewer stuffy noses, for example, is questionable. Despite this, the information obtained from these types of questions provides useful information that fills in the causal chain between policy and changes in adverse health outcomes. Other studies collect more clinical data in smaller samples of workers assessed before and after implementation of a smoke-free policy, although findings may not relate to the general population. For instance, conducting a baseline clinical assessment of a group of nonsmoking bartenders before a smoke-free law is implemented, and then 12-months after the law takes effect, can measure changes in clinical parameters, such as lung function (measured by forced expiratory volume in 1 second and forced vital capacity determinations). These studies are typically expensive to conduct and require clinical facilities. Other studies focus on examining changes in disease rates at the population level in places with and without smoke-free laws, although such studies are rare and it is difficult to identify the independent effect of SHS beyond the effects due to other tobacco control initiatives. The main issues with these studies are that the effect size expected is typically small, effects on nonsmokers specific to the policy cannot be disaggregated from incidental effects on smokers, and it is sometimes difficult to obtain data on the target population of interest. For example, if a large metropolitan area goes smoke-free, but surrounding areas do not, it will be difficult to assess changes in disease patterns, as those who live in the smoke-free metro area may be employed, receive health care, or have other business outside of the city and vice versa. Nationwide policy adoption would limit this concern.

We do not feel that any of these measures is required to be assessed for all smoke-free policy evaluation studies. In the presence of an existing survey already in the field, asking about respiratory effects and related symptoms is encouraged. Clinical or population-based studies examining changes in disease rates are technically demanding studies that require much more planning and resources, and groups with the capacity to conduct these studies are encouraged to do so. Table 5.12 provides a summary of these measures (distal variables).

Incidental outcomes of interest – economic impact, smoking in the home, and smoking cessation

A policy can be thought of as a medication that is intended to treat some condition. If the medication has severe side effects in relation to the benefit it might give the patient, then its utility is diminished. On the other hand, medications can have beneficial side effects making their use even more attractive. In the case of a smoke-free policy, which is the "medication," the key side effect, typically discussed during policy debates, is whether the policy will have an economic impact on businesses. Another incidental effect raised in some policy debates is whether smoke-free policies will encourage smokers to smoke more cigarettes inside their home, thereby increasing their family members' exposure to SHS. Other incidental considerations are whether there are cost savings resulting from increased worker productivity, decreased cleaning costs, and decreased health care costs, and whether the policy increases cessation indicators among smokers. The former two are not discussed in this section, and the latter is described in Section 3.1.

Economic outcomes:

Table 5.13 presents a summary of measures for evaluating the economic impact of a smoke-free policy. Historically, economic

Construct	(a) Self-Reported Changes in Symptoms and Illness
Measure	(Respiratory symptoms) "During the past 4 weeks… a. have you had wheezing or whistling in your chest? (Yes or No) b. have you felt short of breath? (Yes or No) c. do you usually cough first thing in the morning? (Yes or No) d. do you cough at all during the rest of the day or night? (Yes or No) e. do you bring up any phlegm? (Yes or No)" (Sensory symptoms) "In the past 4 weeks… a. have your eyes been red or irritated? (Yes or No) b. have you had a runny nose, sneezing, or nose irritation? (Yes or No) c. have you had a sore or scratchy throat? (Yes or No)"
Sources	Questionnaires; Farrelly *et al.*, 2005a ; Abrams *et al.*, 2006
Validity	Face validity.
Variations	Questions can be adapted to include different conditions.
Comments	These measures do not specifically address the impact of the policy, may be confounded by other factors, and their clinical relevance questioned; however, they do provide a simple way to assess how/why the policy may or may not be working.
Construct	**(b) Clinical Studies Assessing Changes in Worker Health**
Measure	Clinical parameters, such as lung function.
Sources	Clinical exams; Eisner *et al.*, 1998; Allwright *et al.*, 2005
Validity	Clearly valid, but changes in health status could be due to other factors besides the change in SHS exposure.
Variations	Studies can be designed to address health effects in particular subpopulations of interest.
Comments	While these studies provide useful information about the actual near-term health impacts of smoke-free policies, they are costly to perform and require a high level of sophistication to conduct. Implementing this type of study is only recommended for those groups with the resources and research interests to gain a better understanding of exactly how smoke-free policies may change health.

Table 5.12 Health Outcomes in Nonsmokers (Distal Variables or Outcome)

considerations have largely been raised in the restaurant and bar industries, and to a lesser extent, in the tourism and gambling industries. Potential economic impacts in other industries have generally not been studied, nor has there been a call by policymakers for these potential effects to be known.

An ideal economic evaluation would rely on objective measures supplemented with additional measures, such as the self-report of the frequency of visiting bars and restaurants. Objective measures include employment statistics and taxable sales information, as well as statistics on the number of licensed facilities and the number of new and expired licenses. Many countries have established monitoring systems in place that collect these data and access to it is simple; not so for places that do not have such systems. The actual information that can be obtained will differ depending on available data. In the USA, for example, monthly data on the number of employees working in narrow industry subsegments, such as restaurants, can only be obtained from the Bureau of Labor Statistics at the county level. The data are uniformly collected at the national level, are available monthly, and the lag time in reporting the information is a few months, which is relatively quick compared with some taxable sales measures. Taxable sales data share many of the same attributes as employment data; however, tax collection systems are much more variable. Some jurisdictions have a specific tax on meals, which can get tracked independently and is highly specific to that industry segment. Other places rely on general sales or income tax data, which often takes longer to acquire and makes the data less specific to certain industries. Licensure statistics can also provide some insight into the potential economic impact of smoke-free policies. These data only track the number of businesses, so they are not as specific as employment or taxable sales data. They can provide additional complementary evaluation information if available, but relying solely on licensure statistics for an economic evaluation it is not recommended. Both employment and taxable sales measures are excellent objective measures for evaluation, and researchers should investigate what data are available in their country and consider analysing both sets of data.

A useful complement to these objective data sources are self-reported measures of changes in patronage patterns after a policy is implemented. This information can help fill in the causal pathway between a policy and the incidental potential economic losses or gains. In addition, survey questions can be tailored to specific types of venues or to assess more subtle effects. For example, survey questions might assess if people are dining longer, spending more money when going out, or changing the types of places they frequent. These assessments provide a more complete picture of what, if any, economic impact the smoke-free policy is having. The other advantage of reports from individuals is that data can be obtained close to real-time after policy implementation. Employment and taxable sales data take months or years to become available, and then it takes longer still to acquire enough post-law data to establish trends. Policymakers demand an answer to the question of whether the policy has hurt businesses immediately. Survey data, such as described above, can provide an initial glimpse of the potential impact while a case is made that time is needed to examine the objective data sources. Furthermore, self-reported survey items may be the only data source if objective employment and taxable sales data collection systems are not in place.

Another self-reported measure used by some investigators to assess the economic impact of a smoke-free law, is the self-reporting by business owners/managers of changes in sales. Virtually all of the economic studies done that have found an adverse economic impact have utilized this approach, whereas virtually all of the studies based on objective data or individual reports found either no impact or a small positive impact (Scollo *et al.,* 2003). Business owners' lack of support of a smoke-free law has been shown to be associated with more negative fiscal reports, which suggests these data may be

Construct	(a) Changes in Economic Outcomes – Self-reported Consumer Patronage Practices
Measure	Self-report in the change in rate of going out to bars, restaurants, and other locations covered by smoke-free rules. Example question… (Source: ITC Survey) "Do you now visit [pubs/restaurants/etc] more often than [before the law took effect], less often, or about the same amount?" 01 – More Often 02 – Less Often 03 – Same Amount 04 – Don't visit pubs now and/or didn't visit pubs a year ago
Sources	Questionnaires; Hyland & Cummings, 1999a ; Blecher, 2006
Validity	Evidence of utility. In New York City, taxable sales and employment in the hospitality industry increased, while a majority of NYC consumers reported they were dining out the same or more frequently after the 1995 law was implemented (Hyland & Cummings, 1999a; Hyland *et al*., 1999b).
Variations	Questions can be adapted to ask about different locations. Some have obtained more detail by querying about the frequency of going out or actual money spent out, although investigators are cautioned that his information is difficult to recall and obtain from respondents to a population-based survey.
Comments	An economic evaluation should not solely rely on this measure if possible. Ideally, objective measures, like employment statistics or taxable sales data, should provide the basis of an economic evaluation, which can be supplemented with subjective data to help portray a more complete evaluation.
Construct	(b) Changes in Economic Outcomes – Business Owner Self-Reported Change in Sales
Measure	(Source: New York City Restaurateur Survey) Self-reported change in business after a smoke-free regulation takes effect. Example question…"[Over the past two years], would you say your business has increased, decreased, or stayed the same?"
Sources	Questionnaires; Hyland & Cummings, 1999b
Validity	Not recommended as a stand-alone for economic evaluation. Evidence suggests self-report on this item is associated with opinions about the law (i.e. owners who are negative toward smoke-free policies report more negative business outcomes), which may introduce bias into the measurement, and the question is not specific to losses attributable to smoke-free regulations. Some studies show negative outcomes using this approach, but objective data like taxable sales show no impact, or a positive impact, which suggests low validity. This item alone should never be solely relied on for an economic evaluation of a policy.
Variations	Some have obtained more detail, such as asking the actual revenues, but there are considerable levels of missing data when using this approach.
Comments	Many studies do rely solely on this measure; evaluators should be cautioned and aware of the relative merits of this approach.

Table 5.13 Measures for Evaluating the Economic Impact

Construct	(c) Changes in Economic Outcomes – Hospitality Employment Levels
Measure	Government employment statistics for specific industry sectors over time.
Sources	Bureau of Labor Statistics, or other similar government entity; Hyland & Cummings, 1999c
Validity	"Gold standard"
Variations	Can examine specific employment sectors per the policy's specifics.
Comments	This is an excellent measure to evaluate. Study design is enhanced by adding in control employment sectors, as well as data from other jurisdictions not covered by the policy. Comparability of data sources between countries is an issue to consider.
Construct	**(d) Changes in Economic Outcomes – Hospitality Taxable Sales**
Measure	Government tax receipt statistics for specific industry sectors over time.
Sources	Office of Tax and Finance, or other similar government entity; Glantz & Smith, 1994; Hyland *et al.*, 1999b; Cowling & Bond, 2005 ; Blecher, 2006
Validity	"Gold standard"
Variations	Can examine specific sectors per the policy's specifics.
Comments	This is an excellent measure to evaluate. Study design is enhanced by adding in control employment sectors, as well as data from other jurisdictions not covered by the policy. Comparability of data sources between countries is an issue to consider.

Table 5.13 Measures for Evaluating the Economic Impact

biased. For example, in New York City a business owner who was surveyed claimed that losses were experienced (see, for example McLaughlin and Associates Inc (2001) as cited in Scollo *et al.,* 2003), but a review of the objective employment and taxable sales data showed no economic down turn, which means this measure has low validity (see, for example Hyland *et al.*, 1999b and Hyland and Cummings, 1999c, as cited in Scollo *et al.*, 2003). Business owner surveys are excellent for assessing the measures that they used to implement and comply with the new policy, but are not recommended for economic evaluation.

Smoking in the home:

Another potential incidental effect of smoke-free policies is that it may cause smokers, who can no longer smoke at bars for example, to spend more time smoking at home, which leads to greater SHS exposure for other family members. While this issue was not generally raised during policy discussions in the USA, for example, it has gained attention in some European debates. This is a generally understudied area, but it is fairly straightforward to evaluate the likelihood of this potential incidental impact. The simplest approach is to ask smokers how their home smoking

strategy and home smoking behaviour has changed since policy implementation (see Table 5.14). Other approaches could involve tracking how many cigarettes are smoked inside the home before and after the policy in a cohort design. We are aware of only two published studies on this topic. One study used population-based survey data from smokers in four countries, and found that those who lived in a community that implemented a smoke-free bar policy were significantly more likely to implement 100% smoke-free home policies (Borland *et al.*, 2006a). The other study examined differences in smoke-free home policy adoption in Ireland, which had already implemented a smoke-free law, and the UK, which had not implemented smoke-free regulations at the time of the study. It was found that the percent of homes that were smoke-free was comparable between countries, and that Irish smokers consumed fewer alcoholic drinks in the home compared to UK smokers (Hyland *et al.,* 2007). Therefore, this potential incidental effect does not appear to be true, and if anything the opposite, but more studies may be needed.

Smoking cessation outcomes:

Studies have shown that smoke-free worksite policies also increase quit rates and reduce consumption among those who continue to smoke (Fichtenberg & Glantz, 2002a; Fong *et al.*, 2006b). The theorized mechanism of action is that there is a direct impact by decreasing the number of opportunities to smoke and reducing sensory cues for smoking. This reduces the likelihood of relapse during a quit attempt. One large, prospective study on this issue found that smoke-free worksite policies were not associated with a greater rate of trying to stop smoking, but rather quit attempts were significantly more successful (Bauer *et al.,* 2005), which is consistent with what is predicted from the theorized mechanism. Indicators of smoking cessation worth considering for evaluation are quitting, quit attempts, smoking reductions, desire to quit, and utilization of evidenced-based treatments to quit smoking, to name a few. These are described in more detail in Section 3.1 and are not discussed further here.

Moderators

Many important moderating variables are described in Section 3.2. For smoke-free policies, some specific moderating variables of interest include occupation, socioeconomic status, awareness, and beliefs about SHS. Hospitality employees are much more likely to work in an environment where smoking is permitted (Shopland *et al.*, 2004); therefore, a policy that prohibits smoking in the workplace would have a disproportional effect on this population, although we are not aware of studies that have tested this specific hypothesis. Similarly, those with lower socioeconomic status are more likely to work in smoky environments, and should therefore be impacted more by smoke-free policies than white collar workers. Viewing this from a population-perspective, relatively large policy impacts are expected if few workplaces were previously smoke-free and compliance is high; however, there could be little impact if that population is already working in a smoke-free environment. Lastly, those who are aware of smoke-free policies and believe that SHS is harmful are more likely to be compliant with the policy and have lower SHS exposure. Evaluators need to consider moderating variables to best assess how policies may or may not work in population subgroups.

Summary and recommendations

Article 8 of the FCTC calls for governments to increase smoke-free policies at the national and sub-national levels. Evaluating the effects of smoke-free policies is critical to understanding how they work and can be improved. Core constructs to evaluate whether smoke-free policies are working are compliance with the policy and exposure assessment. Based on our assessment of the validity of available data and ease of assessment, we recommend, that in most cases, population-based surveys be used as the primary means for assessing compliance

Measure	**Changes in home smoking rules** (Source: ITC Survey) "Has the smoking ban in public places affected the rules about smoking in your home?" 1 It has made me more strict about the amount I smoke at home when I am with non-smokers. 2 It has made me more strict about the amount I smoke at home in general. 3 It has made me smoke more at home when I am with non-smokers. 4 It has made me smoke more at home in general. 5 It has not affected the rules about smoking in my home. Home smoking policy (Source: GATS) "In your home, is smoking allowed in every place, in some places or at some times, or not allowed in any place?" (Source: Global ATS) "In your home, is smoking allowed in every place, in some places or at some times, or not allowed in any place?" (Source: Adult Tobacco Survey) Car – "Which statement best describes the rules about smoking in your family car or cars? Would you say…Smoking is never allowed in any car, Smoking is allowed some times or in some cars, Smoking is allowed in all cars, or do not have a family car."
Sources	Self-report; Gillespie *et al.*, 2005 ; Borland *et al.*, 2006a
Validity	Face validity.
Variations	Can also assess changes in cigarettes smoked per day in the evening after work, for example, in a cohort design.
Comments	While nationwide SHS policies do not regulate smoking in individual's private homes and property, they may change social norms and increase awareness about SHS harms that may result in individuals implementing such policies on their own. As more workplaces become smoke-free, SHS exposure in the home will be of greater relative importance. This is a relatively understudied area, but has grown to be an important issue in some policy debates.

Table 5.14 Smoking in the Home

with smoke-free policies (Table 5.10). These measures have been validated with ambient air monitoring, as well as biomarkers of exposure. Incidental impacts that may need to be examined are whether smoke-free policies in workplaces affect smoking at home, and how smoke-free policies impact tobacco use behaviour. In some cases, there will be a need to evaluate potential economic impact on businesses, and the use of employment or sales data to assess this impact is recommended.

5.3 Measures to assess the effectiveness of tobacco product regulation

Introduction

Tobacco product regulation is a rapidly emerging area in tobacco control. Scientists, policy makers, and international public health organisations have called for comprehensive regulation of tobacco products with the aim of protecting public health. A handful of countries and jurisdictions have already adopted legislation requiring reporting and testing of tobacco product contents and emissions. Articles 9 and 10 of the WHO Framework Convention on Tobacco Control (FCTC) contain the requirements for regulation of tobacco product contents and emissions, as well as manufacturers' disclosures about the product (Figure 5.6).

As the regulatory landscape evolves around the world, it is essential to evaluate the effectiveness of new regulations and their impact on the product itself and on the population, in order to determine whether regulations are meeting public health goals. The emergence of new legislation and regulatory standards for tobacco products provides a unique opportunity to study changes in the product and in health outcomes over time and across countries and regions. Because product regulations cannot be assessed through randomised clinical trials, researchers and public health officials must employ quasi-experimental designs and utilise opportunities for "natural experiments" through making comparative observations (Fong *et al.*, 2006a). Additionally, it is important to begin collecting baseline data and developing measures and protocols for evaluation, so that the impact of future regulations can be assessed. In 1999, a WHO Conference on the Regulation of Tobacco Products concluded that "The regulatory process must be guided by the best available science and the effects tracked so as to maximize health benefits, minimize unintended consequences, and thereby foster self-correction." (WHO, 2000).

The ultimate test of the impact of a regulation intended to protect public health is to demonstrate a reduction in morbidity or mortality associated with the regulation. However, it can take decades for some effects, such as changes in cancer incidence, to be seen. Thus, measures to assess product regulation have historically focused on the product itself, although such measures have significant limitations for predicting human risk. The need for in-depth product evaluation under actual conditions of use is supported by the history of the development and promotion of "light" cigarettes. Based on standardised machine smoking measurements, the average sales-weighted tar and nicotine yield for US cigarettes decreased by about 70% between the 1950s and 1990s (Hoffman & Hoffman, 2001). Scientists and public health officials initially supported this trend in the 1960s and 1970s (Parascandola, 2005), and it took decades before epidemiologic studies provided definitive evidence that changes in cigarettes designed to lower smoke yields did not in fact lead to any significant decrease in the tobacco-related disease burden (Burns *et al.*, 2001). We now know that much of the apparent decline was due to the use of filter ventilation, which produces markedly reduced machine measured yields, but not necessarily on the amounts smokers actually take in (Kozlowski *et al.*, 1998a).

Laboratory-based product testing remains vitally important, despite its limitations for predicting human risk. First, it supports monitoring of adherence to laws intended to regulate features of product design and performance, such as emission limits based on machine measurements and low ignition propensity laws. Second, it allows for the measurement of differences between products or changes in products that may impact exposure, such as comparing cigarettes that

- **Regulation of the contents of tobacco products.** The Conference of the Parties, in consultation with competent international bodies, shall propose guidelines for testing and measuring the contents and emissions of tobacco products, and for the regulation of these contents and emissions. Each Party shall, where approved by competent national authorities, adopt and implement effective legislative, executive and administrative or other measures for such testing and measuring, and for such regulation.

- **Regulation of tobacco product disclosures**. Each Party shall, in accordance with its national law, adopt and implement effective legislative, executive, administrative or other measures requiring manufacturers and importers of tobacco products to disclose to governmental authorities information about the contents and emissions of tobacco products. Each Party shall further adopt and implement effective measures for public disclosure of information about the toxic constituents of the tobacco products and the emissions that they may produce.

WHO (2003)

Figure 5.6 WHO FCTC Articles 9 and 10: *Regulation of the contents of tobacco products* and *Regulation of tobacco product disclosures,* respectively

heat versus burn tobacco or cigarettes containing tobacco with high versus low tobacco-specific nitrosamine (TSNA) levels. Third, systematic product testing is important because it contributes to the development of general expertise and capacity for tobacco product regulation. Historically, most product-related expertise has been limited to the tobacco industry, and public health scientists have been at a disadvantage in understanding the relevance of product characteristics for health and behaviour, as in the case of "light" and low-tar cigarettes (Parascandola, 2005). While doubtless new, more sophisticated technologies and measures will be developed, such progress will be limited without a network of experienced, public health oriented scientists and technicians.

The task of tobacco product evaluation is complicated by the fact that regulatory requirements are still evolving; for many potential outcomes validated standard measures have not yet been identified. While the FCTC mandates regulation and reporting of tobacco product contents and emissions, guidance for implementation of these articles is still under development by the Conference of the Parties (COP) (http://www.who.int/tobacco/fctc/cop/en/). Thus, it is not clear yet which specific measures will be required in the implementation of the FCTC.

This section will review existing measures relevant to tobacco product regulation as well as discuss challenges and research needs. First, the characteristics of some existing tobacco product regulations will be described to illustrate the range and types of provisions used in current regulations. Second, the section will cover proximal measures for assessing tobacco product regulations, which focus on the product itself. Measures of product content, design and emissions will be discussed, including the limitations of smoking machine protocols for assessing actual human exposure. Third, the section will address distal measures as well, which focus on the impact of regulations for human exposure and risk, including biomarkers and surveillance activities.

Existing tobacco product regulations

Tobacco product regulation remains in its early stages but is evolving rapidly. A number of countries and jurisdictions have adopted product regulations, including ingredient disclosure laws, limits on tar and nicotine yields, low ignition propensity (fire safety) standards, or bans on additives, such as candy flavourings. However, there is little uniformity across jurisdictions in

the content of these laws. Some jurisdictions require constituent disclosure only, while others set standards or limits on content or emissions. Moreover, while some product standards target toxic properties directly (such as by establishing maximum tar or carbon monoxide limits), others target properties that, while not directly harmful, affect addictiveness or consumer appeal (such as by controlling flavour additives that affect the appeal of the product to children).

Currently, there is no centralized, systematic monitoring of tobacco product regulations. The data collected in *Tobacco Control Country Profiles 2003* includes some information on regulation for many countries (Shafey *et al.,* 2003). However, the available data does not specify the details of the regulations (i.e. which constituents are regulated, what product standards or limits are imposed) and it is not updated regularly. As countries continue to debate and enact new tobacco product regulations, there is a need for comprehensive tracking of the evolving regulatory environment.

A few countries and jurisdictions have adopted tobacco product regulations and provide early models of the types of regulatory mechanisms that may be implemented more widely. There are also a number of countries that have adopted International Organization for Standards (ISO) emission limits for tar and nicotine aimed at reducing tobacco related harm, including Brazil, Thailand, China, South Africa, and Malaysia, as well as the European Union (EU).

There are at least five main types of tobacco product regulations that can currently be observed: 1) regulations that require disclosure of product information (such as tar and nicotine content) (Figure 5.7); 2) regulations intended to reduce product toxicity and harm (such as maximum emission limits for tar and nicotine) (Figure 5.8); 3) regulations intended to reduce the addictiveness and/or attractiveness of tobacco products (such as bans on ingredients that impact nicotine delivery or bans on flavour additives that may make a product more attractive to children) (Figure 5.9); 4) regulations intended to prevent fires caused by cigarettes (ignition propensity laws) (Figure 5.10); and 5) bans (or removal of bans) on product categories (Figure 5.11). A few examples are provided in Table 5.15 to illustrate the range of different types of product regulations that are currently being implemented or discussed.

A more detailed presentation of country specific regulations follows:

Canada:

The Tobacco Reporting Regulations, developed under the authority of the 1997 Tobacco Act, require manufacturers and importers of tobacco products to Canada to submit to the Minister of Health information on tobacco product composition and emissions. This includes, for smoked products, information on more than 40 toxic emissions in both mainstream and sidestream smoke under two different smoking regimens, and information on more than 20 specific constituents of whole/unburned tobacco (http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/legislation/reg/index_e.html).

Brazil:

The National Health Surveillance Agency (ANVISA) is charged with regulating a wide variety of consumer products in the interest of public health, including cigarettes and other tobacco products. ANVISA resolution No. 46 (March 21, 2001) establishes maximum tar, nicotine, and carbon monoxide yields for cigarettes, and the tobacco industry is required to submit annual reports that identify and list by brand all ingredients and additives in every tobacco product produced in Brazil (http://www. anvisa.gov.br/eng/tobacco/index.htm).

European Union:

In effect since 2004, a directive of the European Parliament to Member States limits the maximum yield of tar, nicotine, and carbon monoxide in cigarettes manufactured or marketed in the EU (10 mg tar, 1 mg nicotine, and 10 mg carbon monoxide). The directive also requires the tobacco industry to submit to Member States a list of ingredients, and quantities thereof, used in the manufacture of those tobacco products by brand name and type (http://ec.europa.eu/health/ph_determinants

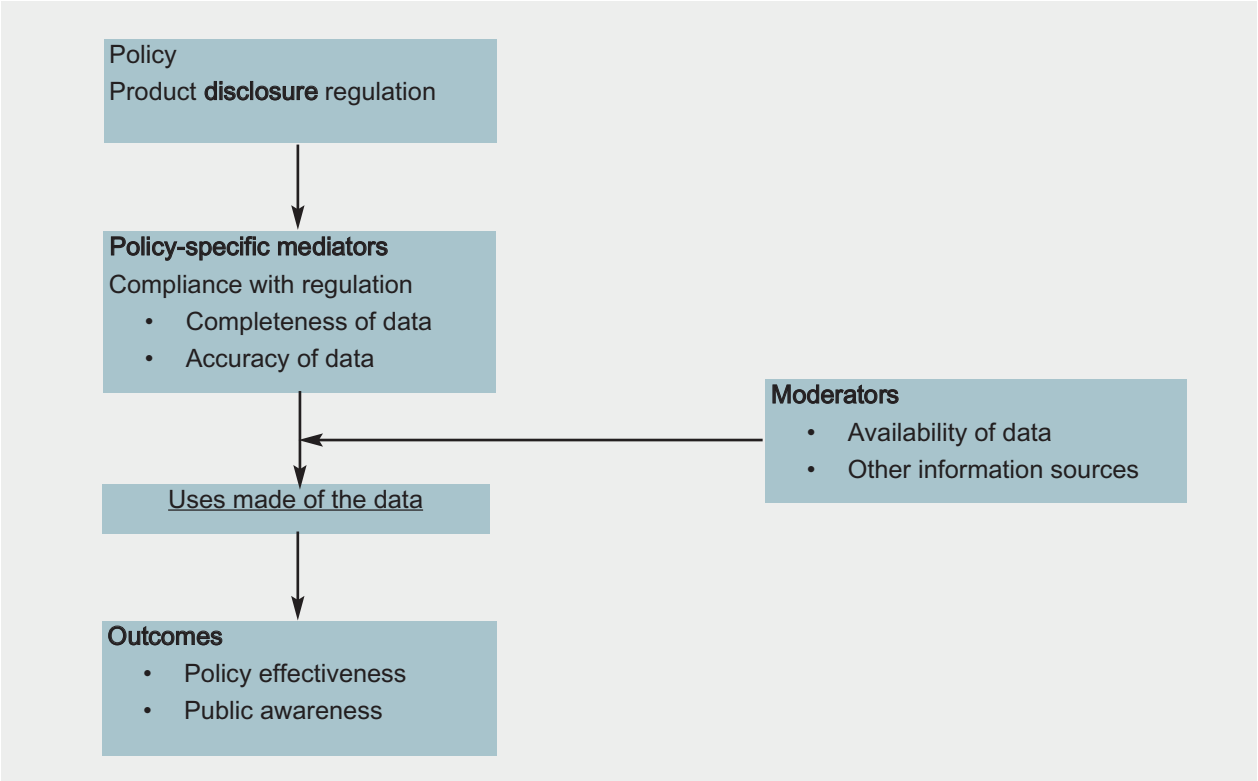


Figure 5.7 Conceptual framework for the evaluation of product disclosure requirements

/life_style/Tobacco/tobacco_en.htm).

United States:

The Comprehensive Smoking Education Act of 1984 and Comprehensive Smokeless Tobacco Health Education Act of 1986 require cigarette and smokeless tobacco manufacturers to submit a list of ingredients added to tobacco to the Secretary of Health and Human Services. However, the law requires that the list not identify the specific brand or company using the ingredients. Smokeless tobacco manufacturers must also report the quantity of nicotine in each product according to standard measures. (Centers for Disease Control and Prevention, 1997a; http://www.cdc.gov/tobacco/FCLA/terms.htm).

Massachusetts:

Manufacturers of cigarettes and smokeless tobacco products sold in Massachusetts must report the product's nicotine yield according to a standardised protocol. The State also proposed a regulation requiring reporting of all ingredients added to cigarettes by brand, but this regulation was barred by a federal court (http://www.mass.gov/dph/mtcp/legal/prodreg.htm).

New York State:

In 2004, New York State became the first jurisdiction in the world to implement reduced ignition propensity (RIP) standards for cigarettes; Canada became the first country to do so in 2005. Both the New York State and Canadian laws stipulate that at least 75% of cigarettes must self-extinguish before burning the full length of their tobacco columns using a

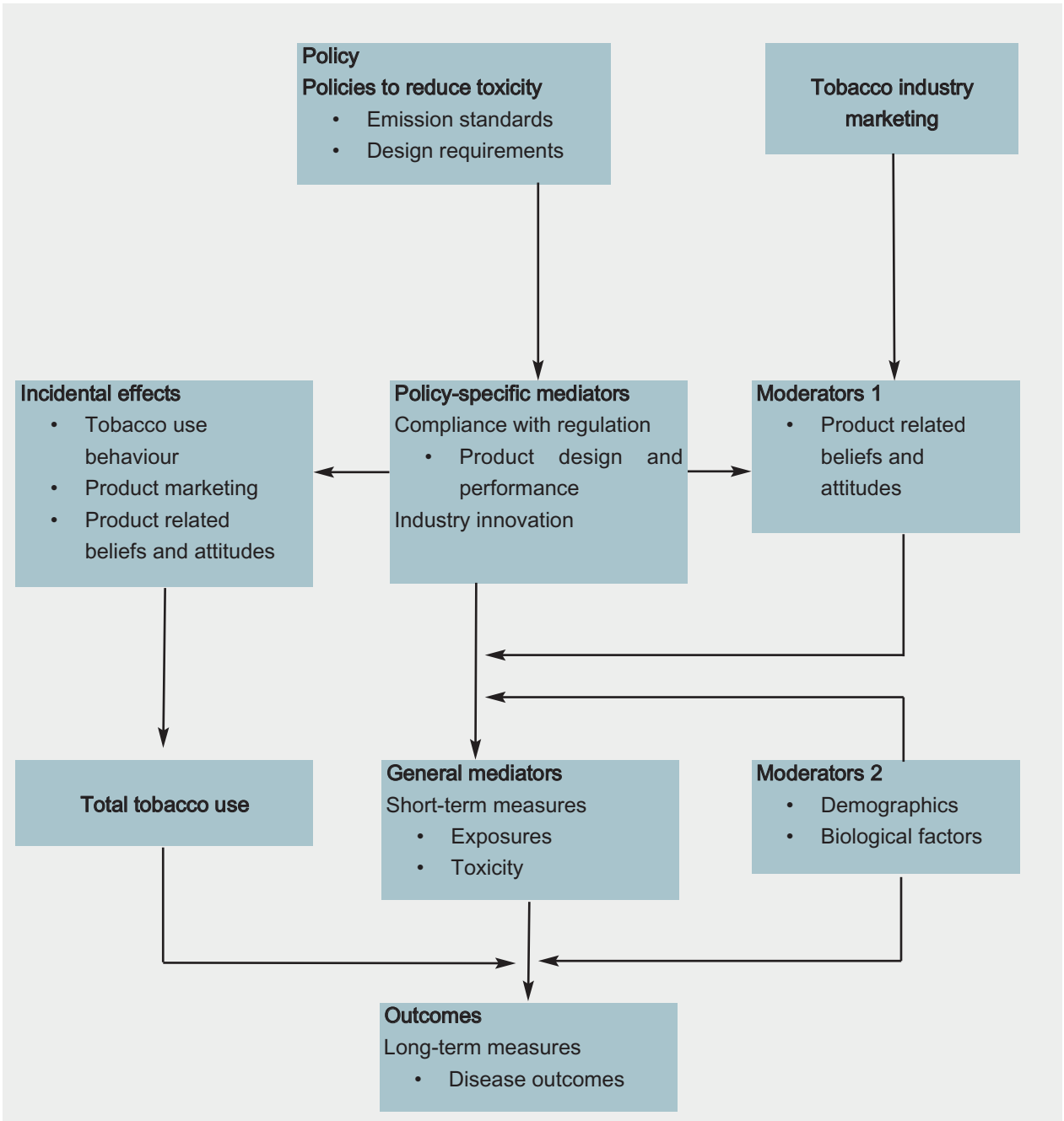


Figure 5.8 Conceptual framework for the evaluation of policies to reduce tobacco toxicity

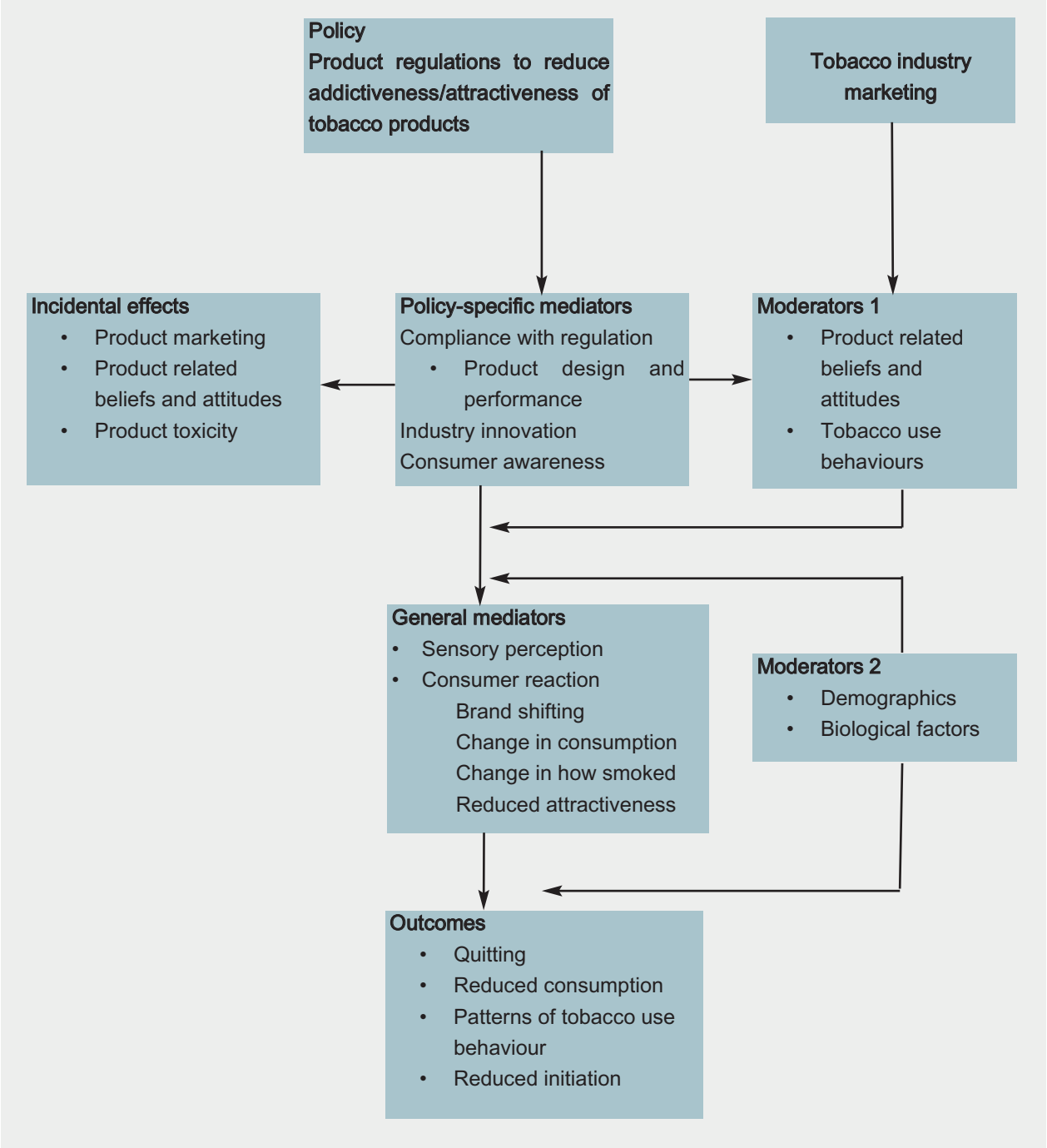


Figure 5.9 Conceptual framework for the evaluation of policies to reduce the attractiveness and/or addictiveness of tobacco products

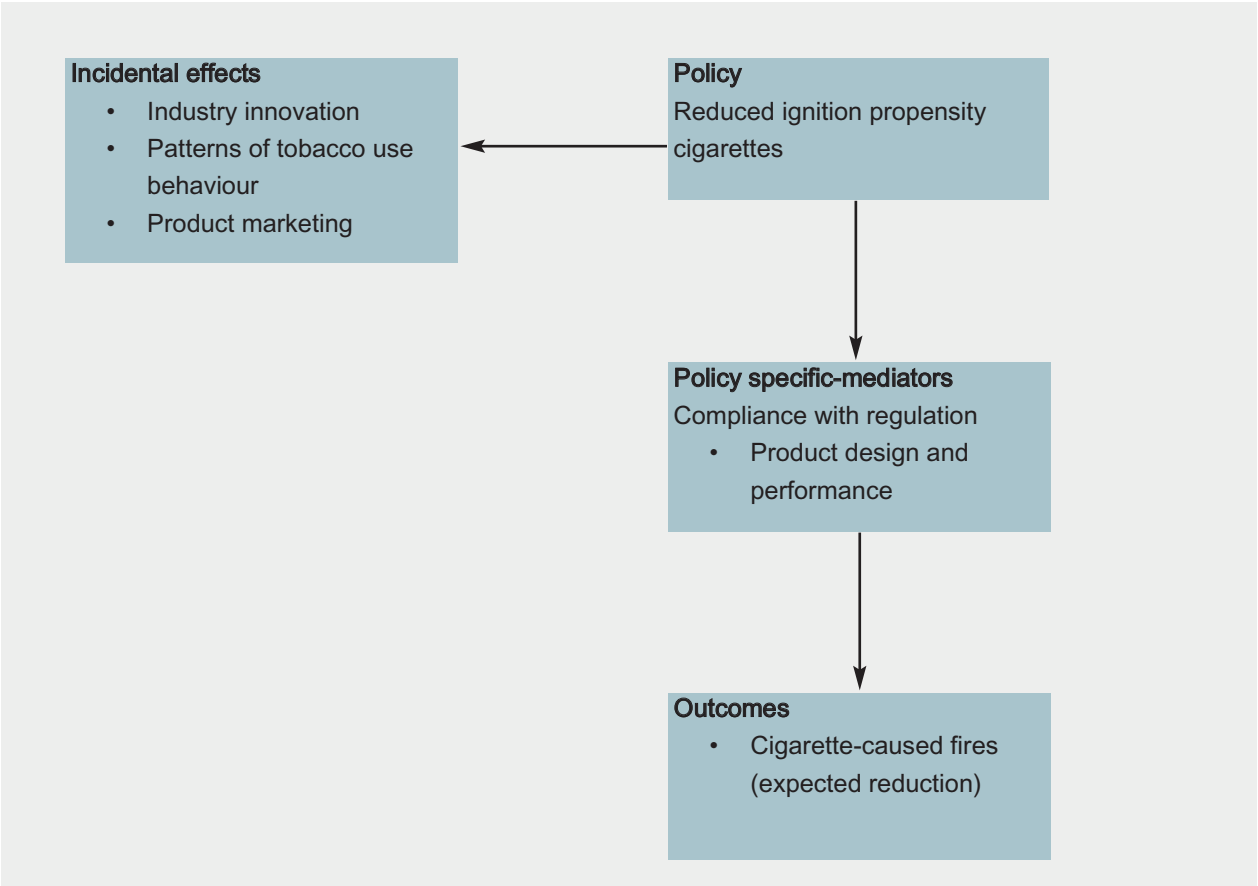


Figure 5.10 Conceptual framework for the evaluation of tobacco product regulation to reduce fires

standardised method for assessing ignition propensity. Both laws use the American Society for Testing and Materials (ASTM) method, which involves positioning a cigarette on one of three standard substrates to generate sufficient heat to continue burning, and thus potentially cause ignition of bedding or upholstered furniture (ASTM E2187-04 *Standard Test Method for Measuring the Ignition Strength of Cigarettes*; http://www.astm.org/cgi-bin/SoftCart.exe/database.cart/redline_pages/e2187.htm?E+mystore).

So far, no jurisdiction has successfully enacted comprehensive regulations governing the design, contents, and emissions of tobacco products. Product performance standards, for example, could be used to reduce known harmful emissions. Currently available data and methods are insufficient to allow for a quantitative estimate of the public health impact of reductions in specific constituents in tobacco smoke. However, evidence shows that there is a wide variation globally between countries and cigarette brands in emissions of tar, nicotine, and carbon monoxide, as well as major carcinogens, suggesting that reductions are feasible and are justifiable on a precautionary basis. A survey of transnational and locally-produced cigarettes in 35 countries found, when measured by a standardised machine

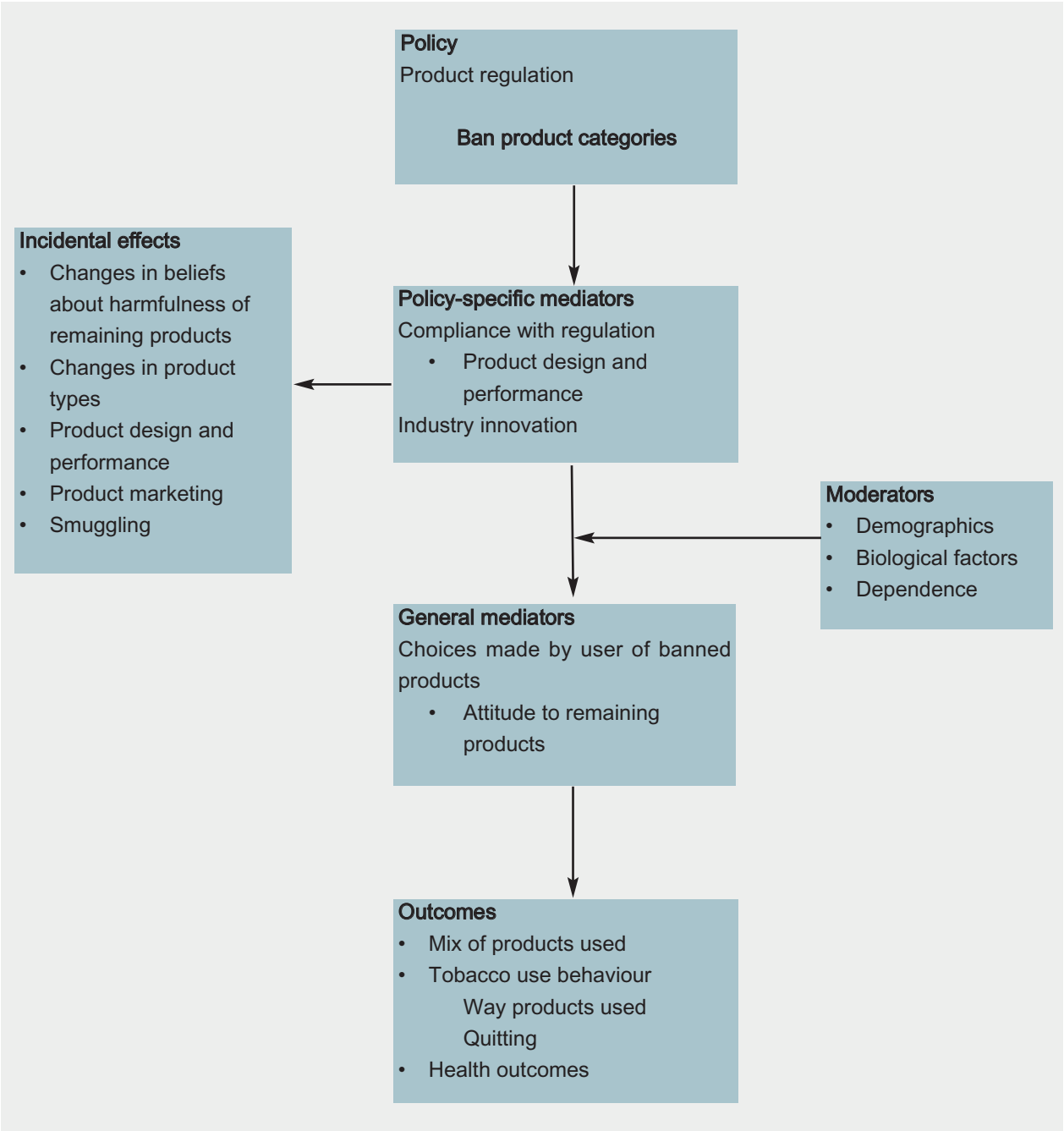


Figure 5.11 Conceptual framework for the evaluation of tobacco product regulation to ban specific product categories

Regulation Type	Requirements
Product Disclosure Example: Canada Tobacco Reporting Regulation	Reporting of 40 constituents in mainstream and sidestream smoke and 20 specific constituents of whole/burned tobacco according to specified protocols.
Reduce Harm Example: European Union Directive 2001/37/EC	Maximum cigarette emission yields: 10 mg tar, 1 mg nicotine, and 10 mg caron monoxide determined by specified machine testing method.
Reduce Addictiveness and/or Attractiveness	Proposed ban on additives that increase the addictiveness of tobacco products.
Reduce Cigarette-Caused Fires Example: New York State Fire Safety Standard for Cigarettes	Mandatory performance standards require that at least 75% of cigarettes must self-extinguish before burning the full length of their tobacco columns, utilizing the American Society for Testing and Materials (ASTM) method for measuring ignition propensity.
Product Bans Example: European Union Directives 2001/37/EC, 92/41/EEC	Prohibits sale and marketing of "all products for oral use, except those intended to be smoked or chewed, make wholly or partly of tobacco."

Table 5.15 Product Regulations

smoking protocol, wide variation in emissions of tar (6.8 to 21.6 mg), nicotine (0.5 to 1.6 mg), and carbon monoxide (5.9 to 17.4 mg), with cigarettes from the Eastern Mediterranean, Southeast Asia, and Western Pacific WHO regions reporting higher deliveries than those from other regions (Calafat *et al.*, 2004). Further analyses from this survey have revealed that mainstream smoke levels of tobacco-specific nitrosamines (TSNAs) and poly-cyclic aromatic hydrocarbons (PAHs) also vary widely across countries, including within the same multinational brand (Wu *et al.*, 2005; Ding *et al.*, 2006). Given the observed variation, one regulatory proposal involves a system for controlling toxins and carcinogens in cigarettes by the establishment of upper limits based on the median of the existing market (Gray & Boyle, 2002). There have also been proposals to reduce nicotine to non-addictive levels to prevent the development of nicotine addiction in young people (Benowitz & Henningfield, 1994). However, so far such proposals have not been implemented in any regulations, and currently there is insufficient evidence to predict what the potential impact of such regulations would be on health outcomes or smoking behaviour.

Regulations requiring tobacco product disclosure, as required in FCTC Article 10, also have an essential role. In order to effectively establish product standards and regulate manufacture of the product, regulators must have valid information about product design, contents, and emissions. Standardised reporting and disclosure by manufacturers assists regulators in monitoring changing trends in product design across the market that may impact public health. Additionally, such disclosures allow for more effective evaluation of the impact and potential unintended effects of new regulatory requirements on product design and emissions.

In order to guide evaluation of tobacco product regulations, it is important to have a conceptual model of the proximal and distal effects of the regulation, taking into account other factors that mediate or moderate those effects (policy-specific mediators, general

mediators, and outcomes). The model should also include other incidental effects of a regulation that are important to evaluating the impact of a regulation on public health. As a general framework, it is likely that the impact of tobacco product regulations on intended health outcomes will be moderated by changes in product design and performance, product marketing, product-related beliefs and attitudes, and tobacco use behaviour, which in turn are expected to influence exposures to tobacco constituents and emissions.

However, because tobacco product regulations can have a range of different goals, multiple conceptual models are needed to understand different types of regulations, just as a variety of methods and measures are needed for evaluating different regulations. Five generalized logic models are provided to guide the development of evaluations of tobacco product regulations (Figures 5.7 to 5.11). These five logic models reflect the five major types of tobacco product regulations identified above (disclosure, reducing product harm, reducing product addictiveness/ attractiveness, preventing fires, product bans). The logic models all start with the introduction of a new policy and then proceed to show a pathway to proximal and distal variables or constructs to be used in assessing the effects of the regulation. Key mediators and moderators, along with incidental effects, are also identified for inclusion in evaluations. For example, a regulation requiring manufacturers to disclose product information to consumers should be evaluated ultimately in terms of its impact on public awareness of the information communicated and the effectiveness of those communications in successfully informing the public about product characteristics. Those effects may be moderated by the availability of relevant data and the presence of other information sources. In contrast, a regulation that aims to reduce product toxicity and harm should be evaluated ultimately in terms of its impact on disease outcomes. Short-term measures of changes in exposures or toxicity, such as use of biomarkers, may substitute for actual measures of disease outcome. These outcomes are likely to be moderated by demographic and biological factors, as well as consumers' product related attitudes and behaviours. Thus, these two types of regulations require very different logic models for their evaluation. Before developing an evaluation plan or protocol, it is important to have a logic model that maps out the goals of the regulation and relevant factors that are expected to influence its effectiveness.

Proximal measures

The most proximal measures of the effectiveness of a tobacco product regulation include measures of the product itself. The first step in evaluating a performance standard regulation, for example, is to measure compliance through product testing. For reduced ignition propensity cigarette laws: does the product meet the full-length burn testing requirements specific in the regulation? For tar and nicotine limits: does the product meet the specified maximum tar and nicotine threshold according to the standardised test method required in the legislation? The specific testing that is required will depend on the requirements and goals of the law. There are a wide variety of product characteristics that could potentially be subject to or be affected by regulation. In addition to assessing compliance, assessment of tobacco product characteristics is important for informing the development of new or modified regulations and for identifying potential unanticipated product changes.

Both tobacco and tobacco smoke are very complex matrices, consisting of thousands of compounds. Over 3044 constituents have been isolated from tobacco (Roberts, 1988); it is estimated that there are over 4800 compounds in mainstream cigarette smoke (Green & Rodgman, 1996). At least 69 carcinogens have been identified in cigarette smoke, including 11 classified as Group 1 known human carcinogens by IARC (Hoffmann & Hoffmann, 1997). Moreover, the composition of cigarettes and cigarette smoke has changed substantially since the 1950s, as the product itself has changed, with changes in tobacco blend, processing techniques, cigarette design, the introduction of filters, and use of additives (Hoffmann & Hoffmann,

1997). At the same time, it is essential to study the product under actual conditions of use, because differences in smoking behaviour can have a substantial influence on product emissions.

However, because of the complexity of tobacco smoke, it is extremely difficult to estimate the health effects of specific constituents in tobacco and tobacco smoke. There have been efforts to quantify the relative contribution to risk of individual tobacco smoke constituents, particularly for cancer, but such estimates are fraught with uncertainty and numerous assumptions. Possibly the most comprehensive such risk assessment, including cancer and non-cancer risk indices for 158 known hazardous chemicals in cigarette smoke, found that these known risk agents underestimated observed cancer rates in cigarette smokers by 5-fold, suggesting that actual exposures were dramatically underestimated and/or that other important carcinogens or mechanisms of action exist that were not included in the risk assessment (Fowles & Dybing, 2003). Further research is needed to understand the individual and combined effects of the many constituents in tobacco and tobacco smoke.

Sampling and preparation

To effectively monitor products as used by consumers, it is essential to follow an effective protocol for obtaining product samples and storing and preparing the product for analysis. Products should be purchased from a range of retail vendors to ensure that the product tested is representative of the product available to consumers and that different manufacturing lot numbers are represented in the sample. In addition, a rigorous protocol should be employed for storing samples. For example, cigarettes and smokeless tobacco should be stored at –70° Celsius in vacuum sealed bags to prevent the effects of aging. Sources of guidelines and protocols for sampling and preparation are available in Table 5.16.

Product content

Official testing of cigarettes has generally focused on measurements of cigarette smoke constituents (i.e. tar, nicotine, and carbon monoxide) using standard machine smoking protocols rather than of the unburned tobacco itself. However, the composition of smoke is directly dependent on the profile of constituents in the tobacco (Fischer *et al.,* 1990). While cigarette design features and human smoking behaviour can dramatically vary the content (both qualitatively and quantitatively) of the smoke and the smoker's exposure, the characteristics of the tobacco are equally important. Moreover, there is wide variation in the concentration of nicotine and other important constituents in the tobacco filter in cigarettes from different brands and countries around the world (IARC, 2004). Additionally, trends in tobacco processing and blending over time may impact public health. For example, while increasing tobacco nitrate levels was seen as a way of reducing

Sampling	**ISO 8243: 2006 Cigarettes: Sampling**
Sample Preparation	ISO 3402: 1999 Tobacco and Tobacco Products: Atmosphere for Conditioning and Testing
	Health Canada: Preparation of Cigarettes from Packaged Leaf Tobacco for Testing (Health Canada, 1999a)
	US Centers for Disease Control and Prevention: Protocol to Measure the Quantity of Nicotine Contained in Smokeless Tobacco Products Manufactured, Imported, or Packaged in the United States (Centers for Disease Control and Prevention, 1997a).

ISO: International Organization for Standardization (www.ISO.org)

Table 5.16 Sampling and Preparations Standards

PAHs in tobacco smoke in the 1960s, in the 1980s scientists recognized that increased nitrate levels were also increasing the yield of nitrosamines in tobacco and smoke (Brunnemann & Hoffmann, 1982; Fischer *et al.*, 1989a). Measurement of constituents in tobacco can provide the earliest point of monitoring for regulation and possible intervention.

There are a range of well established methods for measuring the chemical characteristics of tobacco that have long been in use by tobacco manufacturers and agricultural scientists. Since the 1950s, there have been significant developments in analytical methods for studying tobacco products with the introduction of technologies such as gas chromatography and mass spectrometry (Green & Rodgman, 1996). There are three standard setting organisations that have developed and adopted methods for analysis of tobacco and cigarette smoke: the International Organization for Standardization (ISO), the Association of Analytical Communities International (AOAC), and the Cooperation Center for Scientific Research Relative to Tobacco (CORESTA). The CORESTA board is made up of 14 member companies from the tobacco industry (http://www.coresta.org/Home_Page/Presentation%20of %20CORESTA_April07.pdf).

Additionally, the tobacco-related efforts of ISO have historically been driven primarily by the needs of industry and, thus, they have not adopted methods for many areas of particular interest to public health (i.e. emissions as driven by users behaviour, free-base nicotine, presence of carcinogens) (Bialous & Yach, 2001). Additionally, Health Canada and the US Centers for Disease Control and Prevention (CDC) have published official methods for manufacturer reporting of tobacco constituents. Table 5.17 summarizes the existing methods for whole tobacco analysis and their sources. While an exhaustive discussion of tobacco constituents and associated methods is beyond the scope of this section, a few key agents are discussed here which have particular relevance and importance for product regulation.

Nicotine:

Standardised protocols for extracting and measuring nicotine in whole tobacco using gas chromatographic analysis have been adopted and widely used by industrial and professional organisations (ISO (15152: 2003), CORESTA (No. 62, Feb 2005), AOAC (920.35)), as well as public health agencies (Health Canada, Massachusetts Department of Public Health, CDC). It is important to measure nicotine levels in tobacco as nicotine is the primary driver of smoking behaviour and addiction, and the level of nicotine in tobacco is an essential predictor of nicotine levels in smoke emissions delivered to the tobacco user. A recent report found that nicotine levels in US cigarettes have increased from 1998 to 2005 by about 11%, and concluded that this trend was due primarily to an increase in nicotine in the raw tobacco used in cigarettes (Connolly *et al.*, 2007).

The Massachusetts Department of Health and the CDC also require reporting of the amount of nicotine that is present in the unprotonated, free-base form in smokeless tobacco. This form of nicotine is absorbed more easily through the mucosal membranes in the mouth (Brunnemann & Hoffmann, 1974). Measurements of unprotonated nicotine content in tobacco provide a more accurate assessment of the quantity of nicotine in the product that is delivered to the user (Hoffmann *et al.*, 1995). Free nicotine content in tobacco can be calculated using the Henderson-Hasselbalch equation, which is based on measured pH and nicotine content. This information is important for understanding trends in product use and for providing a basis for monitoring and regulating nicotine content in the product. A CDC study that measured free nicotine in popular brands of smokeless tobacco, found that the brands with the largest amount of unprotonated nicotine also are the most frequently sold (Richter & Spierto, 2003). In smokeless tobacco products, manipulation of tobacco pH and free-base nicotine levels has also been used by the tobacco industry as part of a “graduation strategy,” whereby novice users are introduced to products with lower nicotine

Analyte	Analysis Method	Protocols
Nicotine	Gas chromatographic analysis	Health Canada; CORESTA No. 62, Feb 2005; CDC; AOAC 920.35
Total Moisture	Weight before and after heating in oven at 99° C	CDC; AOAC 966.02
pH	pH meter	Health Canada; CDC
Free Nicotine	Calculated from pH and nicotine using the Henderson-Hasselbalch equation	Centers for Disease Control and Prevention (1997a): Massachusetts Department of Public Health
Nitrosamines	Gas chromatographic analysis	Health Canada; CORESTA (under development); CDC (Song & Ashley, 1999)
Nitrates	Continuous Flow Analysis	Health Canada; CORESTA No. 36, Nov 1994
Metals	Atomic absorption spectroscopy (AAS) analysis	Health Canada; IARC (1986)
Ammonia	High Performance Liquid Chromatography (HPLC)	Health Canada
Humectants	Gas chromatographic analysis	
Pesticide residues	Gas chromatographic analysis	CORESTA No. 2, May 1997; ISO 4389:2000

ISO: http://www.iso.org
CORESTA: http://www.coresta.org/
AOAC: http://eoma.aoac.org/methods
Health Canada: http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/legislation/reg/index_e.html
Massachusetts: http://www.mass.gov/dph/mtcp/legal/prodreg.htm
CDC: US Centers for Disease Control and Prevention: Protocol to Measure the Quantity of Nicotine Contained in Smokeless Tobacco Products Manufactured, Imported, or Packaged in the United States.
Federal Register. Vol. 62, No. 85, Friday, May 2, 1997. p. 24115 - 24117 (recommended method for determination of organochlorine pesticide residues on tobacco)

Table 5.17 Whole Tobacco Analysis Methods

delivery and eventually progress to higher delivery products (Connolly, 1995; Tomar *et al.*, 1995). Thus, continued monitoring of pH levels and free-base nicotine in tobacco is important for monitoring the addiction potential of products (see following subsection on constituents in mainstream and sidestream tobacco smoke).

Nitrosamines:

Tobacco-specific *N*-nitrosamines (TSNAs), *N*-nitrosonornicotine (NNN), 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK), N-nitrosoanatabine (NAT), and *N*-nitrosoanabasine (NAB) are present in both unburned tobacco and tobacco smoke. NNN and NNK play a significant role in cancer induction by tobacco products (Hecht, 1998). The TSNAs are formed from tobacco alkaloids during the curing and

processing of tobacco. Studies have suggested that the tobacco blend may be the most important determinant of TSNAs (UK Laboratory of the Government Chemist, 2000; Harris, 2001). Oriental and flue-cured Virginia tobaccos contain lower levels of nitrate and TSNAs, while higher levels are found in air-cured burley tobaccos (Fischer *et al.*, 1989a; Bush *et al.*, 2001; Peele *et al.*, 2001).

NNN and NNK make a likely target for surveillance and regulation as they play a significant role in tobacco-related cancer, are measurable even in trace quantities, and are specific to tobacco. Moreover, in recent years it has been demonstrated that use of new curing technologies can considerably reduce the levels of TSNA, especially NNK, or even completely eliminate them (Bush *et al.*, 2001; Peele *et al.*, 2001). A study conducted by the CDC comparing TSNA levels in cigarettes purchased in 13 countries and the USA, found that in 11 of the 13 countries locally-purchased Marlboro cigarettes had significantly higher TSNA levels than locally popular non-US brands purchased in the same country (Ashley *et al.*, 2003). Methods for measuring NNN and NNK have been adopted by Health Canada for regulation.

Additives/flavourings:

Additives may include both natural and synthetic agents that impart or enhance flavour. There are hundreds of additives that are used in tobacco products. While in some countries agents may be screened for their direct toxicity, little is known about the fate of these agents after the combustion process. Additionally, additives are used to make tobacco smoke less harsh and to increase nicotine delivery, thus impacting the physiological effects of smoking and resulting behaviours. Ammonium compounds raise the alkalinity of smoke, which increases the level of "free" nicotine delivered to the smoker, and have been employed as an additive in cigarettes (Henningfield *et al.*, 2004). Menthol, a chemical compound which acts as a mild local anesthetic, has been added to cigarettes beginning in the 1920s and 1930s to mask the harshness of tobacco smoke (Reid, 1993).

Detecting flavouring compounds and other additives is complicated by the fact that they may be present in very small quantities and, more importantly, researchers and regulators may lack specific information about their presence. Regulators rely on information from annual reports of additives used and their quantities by cigarette brand, such as in the EU, but many countries do not yet have such requirements. Because of the hundreds of additives that may be in use, testing for many of them is impractical. At least one study has quantified the presence of 12 potentially toxic flavour-related compounds in cigarette tobacco, including coumarine and safrole, and found that 62% of 68 brands tested contained one or more of these 12 compounds (Stanfill & Ashley, 2000). The UK Department of Health maintains a list of permitted additives to tobacco products (now numbering over 600) along with maximum inclusion limits, although their effects after combustion have generally not been tested (http://www.advisorybodies.doh.gov.uk/scoth/technicaladvisorygroup/additiveslist.pdf).

Evaluation of the impact of product regulations that control additives is limited by inadequate information and scientific data about the presence of additives in products by brand, and their potential effects on behaviour and health outcomes.

Product design

Cigarette design has evolved over the past half century, with the introduction of filters, changes in tobacco processing techniques, and the introduction of new materials and technologies. The resulting changes in product design and characteristics can have a substantial impact on the exposure a smoker receives. The types of materials used in filters and filter design can alter the chemical composition of the smoke that is inhaled, including the levels of carbon monoxide and other harmful constituents. Additionally, use of expanded or reconstituted tobacco in cigarettes can affect tar and nicotine yields and the profile of constituents. Cigarette length, circumference, and packing density can also alter the chemical composition of the smoke (Hoffmann & Hoffmann,

1997). Specific design features have also been employed to reduce cigarette ignition propensity, such as reduced tobacco density, reduced paper porosity, decreased circumference, and the removal or reduction of burn additives.

Physical characteristics of tobacco products should be measured in order to inform the development and implementation of tobacco product regulations and to support evaluation of regulations. The WHO Study Group on Tobacco Product Regulation (TobReg) has provided a recommended list for product characteristics to be reported by manufacturers for all brands on an annual basis (WHO Study Group on Tobacco Product Regulation, 2004; http://www.who.int/tobacco/global_interaction/tobreg/goa_2003_principles/en/index.html).

Table 5.18 includes the TobReg recommendations and additional product characteristics that should be measured to assess the impact of regulation on product design; reference numbers are provided for official laboratory protocols where they exist. This list is not exhaustive and should be revised regularly to account for new types of products and design innovations, such as new potential reduced exposure products (PREPs) that employ unconventional technology. These product characteristics are not necessarily direct targets of regulation or indicators of effectiveness of regulations in all cases. They should be considered, however, as useful measures for supporting the development and implementation of regulations, such as by revealing unexpected product changes in response to regulations (see following section on ventilation). Because most of the measures are routinely used by manufacturers in product characterization and quality control, such information should be requested from manufacturers by regulators where possible.

Cigarette ventilation:

Since the 1960s, cigarette filter ventilation has been the dominant design feature employed by manufacturers to reduce machine measured tar and nicotine yields (Kozlowski *et al.*, 2006). Small pinholes on cigarette filters allow

Product Characteristics	Measurements
Raw Materials	Tobacco blend, weight of tobacco, percentage of reconstituted tobacco, percentage of expanded tobacco, moisture content, firmness, contaminants (i.e. glass, pesticides, heavy metals).
Filter	Type, length, weight, density, ventilation[a], draw resistance[b], fiber residues, charcoal content.
Cigarette Body	Rod length, tipping paper length, diameter[c], air permeability[d].
Emission	Aerosol particle size with and without filter.
Ignition Propensity	Percent self-extinguishing.

[a]ISO 9512: 2002 Cigarettes - Determination of ventilation - Definitions and measurement principles
[b]ISO 6565: 2002 Tobacco and tobacco products - Draw resistance of cigarettes and pressure drop of filter rods - Standard conditions and measurement
[c]ISO 2971: 1998 Cigarettes and filter rods - Determination of nominal diameter - Method using a laser beam measuring apparatus
[d]ISO 2965: 1997 Materials used as cigarette papers, filter plug wrap and filter joining paper, including materials having an oriented permeable zone - Determination of air permeability

Table 5.18 Product Characteristics to be Measured to Assess Impact of Product Regulation

the smoke to be diluted by air drawn in by the smoker. However, studies have shown that smokers tend to place their fingers over these vent holes in order to derive a desired level of nicotine (Kozlowski *et al.,* 1980). Additionally, smokers puff harder to compensate and the greater flow through the cylinder also reduces the proportion of air that comes in through the vent holes. Because of this flexibility in the cigarette design, machine measured ISO/FTC tar yields do not reflect the actual range of exposures smokers receive. A study comparing ventilation (measured as the percentage of air drawn through the filter vents) across 32 brands of US cigarettes, with FTC tar yields ranging from 1 mg to 18 mg, found that the degree of ventilation (from 0 to 83%) varied inversely with standard tar, nicotine, and CO yields, suggesting that ventilation is a key determinant of machine measured yields (Centers for Disease Control and Prevention, 1997b). Similarly, another study accounted for 95% of the variance in ISO measured levels as a function of extent of filter venting (King & Borland, 2004).

A recent study assessed how UK cigarette manufacturers modified their products in order to comply with the EC 10-1-10 maximum yield regulation. Comparing 10 cigarette brands before and after the regulation was imposed, they found that machine measured tar was reduced from 11-13 mg to 10 mg for each brand, carbon monoxide yields dropped significantly from a median of 13 to 10 mg, as well as nicotine from a median of 1.0 mg to 0.9 mg. However, the only product design feature that showed consistent change was the amount of filter ventilation, as the median increased by 479% from 1999 to 2005. In contrast, other product design characteristics that were measured in the study, including filter weight, filter length, and tobacco length, showed no changes (O'Connor *et al*., 2006a). This study illustrates the importance of monitoring product design over time against a baseline level to understand how products are modified in response to new regulations, and whether the public health objectives of the regulation are being met. An alternative proposal involves imposing maximum tar, nicotine, and carbon monoxide yields along with a ban on filter vents (Kozlowski & O'Connor, 2002; Kozlowski *et al*., 2006).

Amount of ventilation should be measured in cigarettes, particularly for evaluating the introduction of new regulatory limits on emissions. Additionally, given the elasticity in exposures from ventilated cigarettes, measurements of emissions should take this variability into account, such as by measuring emissions in relation to a fixed amount of nicotine or per milligram of nicotine.

Reduced ignition propensity:

Reduced ignition propensity (RIP) regulations are relatively new, so limited data is available on their impact and effectiveness. One study conducted to evaluate the impact of the New York law, found that the average percentage of full-length burns was 10% for five leading brands sold in New York after the law went into effect, compared with 99.8% for cigarette brands from California and Massachusetts (Connolly *et al*., 2005). These findings confirm that the law did result in changes to the product design that achieved the aims of the legislation. Product testing can be used to assess compliance and product performance following the introduction of RIP laws. It is also important to evaluate smokers' reactions to changes in cigarette design to identify potential unintended effects on smoking behaviour. A survey of adult smokers' reactions to RIP cigarettes found that while smokers in New York State were more likely to report that their cigarettes went out between puffs, they were no more likely than smokers in states without RIP laws to report differences in cigarette taste, suggesting that RIP cigarette laws do not substantially impact consumer acceptability (O'Connor *et al*., 2006b). Moreover, proximal measures of the product itself cannot assess more distal outcomes, such as changes in the number of fires caused by cigarettes. Distal measures and surveillance are discussed in the following section.

Product Emissions

Measuring the contents and characteristics of tobacco smoke has been the primary focus of

tobacco product testing and regulation efforts since the 1960s. Measuring the contents of tobacco smoke provides direct information about the agents the smoker is exposed to. However, these measures also have substantial limitations; while they allow for the identification of important constituents in tobacco smoke, they do not necessarily reflect exposure under actual smoking conditions. Measurements of product emissions have typically relied on machine collection of tobacco smoke, which does not reflect actual human smoking behaviour. This section will review various machine smoking protocols, and their limitations, and will then discuss specific constituents in tobacco smoke that have been proposed for surveillance and regulation.

Machine smoking methods:

Machine smoking methods for measuring tar, nicotine, carbon monoxide, and other constituents in cigarette smoke have been widely used in many countries over the past 30 years. The procedure involves having a machine "smoke" cigarettes according to fixed parameters that determine the frequency, duration, and volume of puffs, as well as the butt length. The particulate matter is collected onto a Cambridge filter pad made of extremely fine diameter glass fibers. Mainstream smoke particulates are collected on filter pads located behind the cigarette port, while sidestream smoke is collected with the use of BAT "fishtail" devices, which allow smoke from the end of the cigarette to travel up a glass enclosure to a filter pad located at the top. Filter pads are weighed before and after a "smoking" run to determine the Total Particulate Matter (TPM) (the amount of particulates accumulated on the filter pad). A solvent is used to remove the chemicals from the filter pads, and once this extraction is complete, various chemical and physical separation techniques are used to isolate the desired component(s). Once the desired chemical has been isolated, various analytical methods (such as gas chromatography with mass spectrometry) are used to determine the amount of chemical collected. Gas phase chemicals, such as carbon monoxide, may pass through the filter pads and into collection bags for measurement.

To ensure consistency across measurements, standard parameters are used to control the machine's puffing activity. The parameters most widely in use were based on a protocol outlined by the US Department of Agriculture (Ogg, 1964); a similar protocol had been proposed by American Tobacco Company researchers in 1936 (Bradford *et al.*, 1936). The protocol called for 2-second, 35-mL puffs to be taken until a 23-mm butt length remained on the cigarette. These parameters were somewhat arbitrarily selected based on informal observations; Ogg reportedly stated that he arrived at the parameters he chose by informally observing people smoking, timing them with the aid of a stopwatch, and measuring the length of the "unsmoked" cigarette left in the ashtray (Harold & Pillsbury, 1996). When the US Federal Trade Commission adopted this method for use in its testing laboratory, the agency acknowledged that these parameters were not intended to mimic the smoking behaviour of any particular individual or even an "average" smoker, but the application of a uniform standard would, they stated, allow for meaningful comparisons across products (Press release, August 1, 1967). ISO adopted a similar set of parameters in their cigarette testing method (ISO Standard 3308: 2000 (4th edition), *Routine Analytical Cigarette-Smoking Machine: Definitions and Standard Conditions*).

However, beginning in the 1980s, a more profound understanding of smoking behaviour revealed that smokers who switched to cigarettes with lower machine measured tar and nicotine yields modified their smoking behaviour to compensate by taking more frequent puffs, inhaling the smoke more deeply, covering up filter ventilation holes, and smoking more of each cigarette (Benowitz *et al.*, 1983; National Cancer Institute, 2001). More accurate measures of the actual smoke exposure of a given individual can be obtained through the study of smoking topography, where the smoker uses a mouthpiece connected to a device that measures parameters of smoking behaviour (such as number of puffs, puff volume, duration, velocity, and the intervals between puffs) (Djordjevic *et al.*, 2000; Lee

et al., 2003). However, while smoking topography measurements are valid for assessing individual exposure, the parameters vary widely across the population and no single set of smoking parameters can effectively represent this variation.

Because of growing concerns about the validity of the FTC/ISO parameters, alternative machine smoking regimens have been proposed. In particular, the FTC and ISO smoking regimens do not account for the fact that smokers may cover ventilation holes with their fingers, and alternative smoking regimens have attempted to address this. The Commonwealth of Massachusetts in the USA currently tests cigarettes with a 45 mL puff drawn twice per minute with 50% of the filter vent holes blocked (Commonwealth of Massachusetts, 2007). Canadian government testing standards require a more intensive smoking regimen, where 55 mL puffs are drawn twice per minute with 100% of the vent holes blocked (Health Canada, 1999b). While these regimens also cannot represent the wide variation in human smoking patterns, they may be less likely to underestimate actual human exposure by using more intense puffing parameters. This may be especially important for lower yield products for which smokers may compensate with more intense puffing behaviour.

A "compensatory" machine smoking regimen was proposed; rather than smoking all brands using the same puffing regimen, the compensatory regimen attempts to mimic the systematic differences in human smoking across different products, whereby lower nicotine yield brands are smoked more intensely. It was suggested that the puff volume and puff frequency be varied according to the ISO nicotine yield. For brands with <10 mg tar, a 40 mL puff is taken every 60 seconds. With every decrease of 0.1 mg nicotine, the puff volume rises by 4 mL and the puff frequency falls by 4 seconds. For example, a cigarette with 0.5 mg nicotine under the ISO method would be smoked at 60 mL puffs every 40 seconds, whereas a 0.1 mg cigarette would be smoked at 76 mL puffs every 24 seconds (Kozlowski & O'Connor, 2000). Another alternative is to tie analysis of constituents to a fixed nicotine level whereby cigarettes are smoked to predetermined nicotine yields and the levels of other constituents assessed from that (Hammond *et al.*, 2007b). Alternatively, TobReg of WHO has recommended use of yields per mg of nicotine, using standard puffing regimens. (WHO Study Group on Tobacco Product Regulation, 2004).

A recent study compared the performance of these four smoking regimens against actual human smoking patterns and biological measures of exposure to assess how well they reflect actual exposures smokers are likely to receive (Table 5.19) (Hammond *et al.*, 2006b). The aim of the study was to compare measures of smoke volume and nicotine uptake among human smokers against the puffing variables and nicotine yields generated by the four smoking regimens, as well as a Human Mimic regimen where brands were machine smoked using puffing behaviour recorded from human smokers in the study. Participants in the study smoked cigarettes through a portable smoking topography device to record their smoking behaviour, and they also provided saliva samples to be analyzed for cotinine. The study found that, using the Human Mimic condition as a benchmark, subjects were exposed to tar, nicotine, and carbon monoxide levels that were 2 to 4 times greater than the ISO yields, suggesting that the ISO standard seriously underestimates actual human exposure. Moreover, while the Canadian intense smoking conditions are considered to represent the maximum emissions to which a smoker is likely to be exposed, the study found that total smoke volume was not significantly different from the actual smoke volume as measured in the participants when smoking their usual brand. Among those subjects who were experimentally switched to a lower yield brand, all four smoking regimens produced a lower volume of smoke than the Human Mimic. Comparing these findings to the measured salivary cotinine levels further reveals the limitations of machine smoking methods. The yields from the Massachusetts, Canadian, and Compensatory regimens were no better at predicting measures of

	FTC	ISO	Massachusetts	Canadian	Compensatory
Puff Volume (mL)	35	35	45	55	40
Puff Duration (seconds)	2	2	2	2	2
Interpuff Interval (seconds)	60	60	30	30	30
Ventilation Hole Blockage (%)	0	0	50	100	50
Butt Length	23 mm or filter + 3 mm	Filter length + 8 mm or filter over-wrap+ 3mm	Filter length + 8 mm or filter overwrap + 3 mm	Filter length + 8 mm or filter overwrap + 3 mm	Filter length + 8 mm or filter overwap + 3 mm

Adapted from Hammond *et al.*, (2006b)

Table 5.19 Recommended Machine-Smoking Regimes for Cigarette Testing

nicotine uptake than the ISO yields. Even the Human Mimic condition was only moderately correlated with salivary cotinine levels, reflecting the wide variability in uptake based on nicotine metabolism among smokers even when smoking the same brand. A subsequent study comparing emissions data from 238 Canadian cigarette brands tested under ISO and "Canadian intense" machine smoking conditions, found that the more intense protocol was not necessarily more representative of actual human smoking behaviour and exposure (Hammond *et al.*, 2007b).

Standardised machine testing regimens lack validity as measures of actual human exposure. Despite its limitations, however, machine testing using ISO and alternative parameters remains valuable for informing the development and implementation of product regulations and, where relevant, for measuring basic compliance with constituent limits based on standardised machine testing regimens. The WHO Study Group on Tobacco Product Regulation (TobReg) has recommended that standardised machine smoking tests be used by scientists and regulators "to the extent that it provides a basis for a comparison of the results with new testing protocols until protocols that reflect variations in human smoking behaviour according to different cigarette designs are developed." (WHO Study Group on Tobacco Product Regulation, 2004). Despite its limitations for predicting actual human exposures, machine testing can provide important information on cigarette engineering and how differences in cigarette design may affect smoke emissions.

There remains a need for further development of methods for collecting smoke emissions that are more representative of actual human smoking exposures. Additionally, some promising approaches to account for variations in smoking behaviour based on nicotine titration warrant further development, including measurement of constituent yields per milligram of nicotine and analysis of cigarette filter stains (Strasser *et al.*, 2006)

Constituents in mainstream and sidestream tobacco smoke

Mainstream cigarette smoke is a complex and dynamic mixture of thousands of constituents that are distributed between a vapour phase and a particulate phase (Jenkins *et al.*, 2000). Since the 1950s, following the first epidemiologic studies linking smoking and lung cancer, dozens of

carcinogens and other harmful constituents have been identified in tobacco smoke. The primary focus has been on PAHs, such as benzo [*a*]pyrene and TSNAs, such as NNK, which are considered to be major lung carcinogens (Hecht, 1999). Carbon monoxide in cigarette smoke has also been extensively studied and is likely to contribute to atherosclerosis, and other cardiovascular diseases, by reducing delivery of oxygen through the body (US Department of Health and Human Services, 2004). It is not possible to discuss the significance of each constituent in this section, but a thorough list of major toxic and carcinogenic constituents in the vapour phase and particulate matter of cigarette smoke is provided in IARC Monograph 83 (IARC, 2004). The WHO TobReg study group has developed a recommended list of constituents to be reported or measured in mainstream and sidestream smoke (2004). Additionally, Health Canada requires manufacturers to report more than 40 specific constituents annually for each brand in both mainstream and sidestream smoke. Though essentially the same list of constituents is measured for both mainstream and sidestream smoke, it is important to do measurements for both types of emissions because their quantities may differ. These constituents are listed in Table 5.20.

A few compounds believed to be particularly important are briefly discussed here:

Nicotine:

Measuring nicotine emissions is central to evaluating the addictive potential of tobacco products. Standardised methods for measuring nicotine in machine collected smoke have been widely used, but their ability to predict actual nicotine intake is restricted by the limitations of standardised machine smoking parameters. It is also important to measure the proportion of nicotine that is available in the unprotonated, free-base form, which is more easily absorbed by the body. Research has shown that levels of free-base nicotine vary substantially across different types of tobacco and tobacco product brands, and that the tobacco industry has manipulated the free-nicotine content of tobacco products through additives, such as ammonia (Ferris *et al.*, 2006). A laboratory smoking device and a gas chromatograph-mass spectrometer were used to measure the amount of free-base nicotine in the particulate matter of mainstream cigarette smoke, and found that significant amounts of nicotine in the particulate matter can be in free-base form (Pankow *et al.*, 2003). Similarly, a research group from the CDC found that the measured ranges of free-base nicotine in smoke particulate matter were remarkably similar over the different tar and nicotine delivery categories of full-flavoured, light, and ultra-light cigarette brands, sug-gesting that standard tar and nicotine yields do not provide a valid estimate of actual nicotine emissions (Watson *et al.*, 2004b).

Polycyclic aromatic hydrocarbons:

Polycyclic aromatic hydrocarbons (PAHs) are a diverse group of carcinogens formed during the incomplete combustion of organic material, such as tobacco. They are found in tobacco smoke, broiled foods, and in occupational settings, such as iron and steel foundries. Benzo[*a*]pyrene is the best known member of this class of compounds and has been classified by an IARC expert panel as “carcinogenic to humans” (Straif *et al.*, 2005).

N-Nitrosamines:

Tobacco smoke nitrosamines (TSNAs) include a large group of carcinogens that are known to induce tumours in a variety of animal species. TSNAs, such as NNN and NNK, are chemically related to nicotine and nornicotine, a secondary amine tobacco alkaloid, and are thus only found in tobacco products. An IARC working group on smokeless tobacco and tobacco-related nitrosamines concluded that exposure to NNN and NNK is “carcinogenic to humans” (Cogliano *et al.*, 2004).

Aromatic amines:

Aromatic amines were first identified as carcinogens in workers in the dye industry. Of these, 4-aminobiphemyl and 2-naphthylamine are well-established human bladder carcinogens (IARC, 1987).

The 1999 Massachusetts Benchmark Study provided the

Health Canada	TobReg Mimimum
Nitrosamines	NNN, NNK, NAT, NAB
Acrylonitrile	
3, 4 Aminobiphenyl	
1,2 Aminonaphthalene	
Ammonia	
Arsenic	Arsenic
Benzene	
Benzo[*a*]pyrene	
1,3-Butadiene	
Cadmium	Cadmium
Carbonyls	
Chromium	Chromium
Eugenol	
	Formeldahyde
Hydrogen Cyanide	Hydrogen Cyanide
Isoprene	
Lead	Lead
Mercury	Mercury
Nickel	Nickel
Nitrogen Oxides	Nitrogen Oxides
Phenolics	
Pyridine	
Quinoline	
Selenium	Selenium
Styrene	
Toluene	
Filter efficiency	
pH	
Tar, nicotine, carbon monoxide	Tar, nicotine/free nicotine, carbon monoxide
	Ratio of nicotine-free dry particulate matter to nicotine yield

Table 5.20 Emissions Candidates for Surveillance

most comprehensive data to date on the profile of smoke emissions of contemporary cigarettes. Eighteen leading cigarette brands from the USA delivering a range of tar values (from 1 mg to 26 mg per cigarette according to FTC parameters) were screened for 44 constituents using both the FTC and Massachusetts machine smoking methods. The primary constituents varied dramatically across the brands, including total tar (6.1 mg to 48.7 mg per cigarette), carbon monoxide (11.0 mg to 40.7 mg per cigarette), and nicotine (0.50 mg to 3.32 mg per cigarette) (Borgerding *et al.*, 2000; IARC, 2004). The study also illustrated the limitations of ISO tar and nicotine yields for predicting doses of specific toxins and carcinogens in tobacco smoke. One analysis of the Benchmark data showed that FTC tar, nicotine, and carbon monoxide yields were poor predictors of TSNA

yield per cigarette, suggesting that information about the tobacco blend could be more informative for predicting TSNA emissions (Harris, 2001).

Indeed, measured yields of constituents can vary substantially depending on the smoking parameters used for machine measurements. One analysis found that the yields of six IARC Group I carcinogens (benzene, cadmium, 2-aminonaphthalene, nickel, chromium, and 4-aminobiphenyl) in mainstream smoke, were an average of 2-4 times higher when measured by the more intense Health Canada parameters than by ISO parameters (IARC, 2004). Another study of mainstream smoke from three popular brands of US cigarettes purchased on the open market in 29 countries worldwide, showed little variation in tar and nicotine, but substantial differences in the yields of NNN and NNK within each brand (Gray *et al.,* 2000). Additionally, analyses have shown that blocking filter ventilation holes can alter the characteristics of mainstream smoke, including increasing the delivered doses of specific carcinogens and hazardous agents (Brunnemann *et al.*, 1990). These analyses suggest that standard ISO machine measured tar and nicotine ratings cannot be relied upon to estimate emissions of toxic constituents. Further research is needed to understand how varying smoking parameters may affect the contents of cigarette smoke.

Cigarette smoke is also highly dynamic, and the profile of smoke constituents varies over time and across puffs in response to changes in temperature and dilution of smoke and other factors. The distribution of individual constituents across the particulate and gas phases of smoke also changes over time; volatile and semi-volatile compounds, such as benzene and 1,3-butadiene, can be present in significant quantities in both phases. Recently, a high throughput method for analyzing volatile organic compounds in smoke was published (Polzin *et al.*, 2007). However, measuring this dynamic mix in real-time to determine how exposure varies over a series of puffs, for example, is extremely complex. Efforts have been made to characterize volatile compounds in smoke in real-time using time-of-flight mass spectrometry, but this application is experimental and requires state of the art equipment (Adam *et al.,* 2006).

Distal measures

Biological Impact:

The ultimate test of the success of tobacco product regulations in protecting public health would be to observe actual reductions in tobacco-related disease incidence. Population level trends in lung cancer incidence, for example, have reflected changes in cigarette smoking over time. However, such long-term health outcomes do not represent an effective target for regulation, because of the delay between exposure and the appearance of disease symptoms, which can, as with cancer, take decades.

Biomarkers of exposure and biological impact show substantial promise for assessing early effects of tobacco use that are relevant for later disease outcomes. Disease risk is presumed to be a function of the amount, site, and duration of the exposures. Thus, biomarkers of exposure may provide more accurate prediction of disease outcomes than standard measures of tobacco consumption. In particular, there are substantial differences in how individuals use tobacco products, and how their bodies respond to chemical agents in tobacco smoke that are not reflected by simply measuring number of cigarettes per day or use of standardised machine smoking to predict exposures. Additionally, biomarkers may play a particularly important role in the assessment of how differences between products or changes in product design or constituents may impact health. For example, biomarkers of toxic effects or biological damage can provide early indications of the impact of potential reduced exposure products or constituent limits on disease outcomes.

Biomarkers can be divided into at least two major categories (Hatsukami *et al.*, 2006):

- Biomarkers of internal exposure: biomarkers that provide a direct or indirect measure of the quantity of a tobacco-derived constituent or

metabolite in the body. These will not always be closely related to intake because of differences in rates of metabolism.

- Biomarkers of potential harm: biomarkers that measure a biological effect or binding of a tobacco constituent or metabolite in a target organ or tissue. For example, carcinogen-DNA adducts can be used to measure the presence and activity of a specific carcinogen in target tissue. Further along, this also includes biomarkers that measure actual damage to organs or tissues, such as genetic mutations or chromosomal aberrations, which may or may not lead to disease.

It is important to distinguish between biomarkers of exposure versus biomarkers of biologic effects or disease; it may be possible to show a reduction in exposure while the impact on disease outcomes remains uncertain. Additionally, it is helpful to distinguish between biomarkers specific to a particular chemical, such as NNAL, and biomarkers that assess the impact of complex exposures, such as urine mutagenicity.

With the rise of genomics and advances in molecular biology the field of cancer-related biomarker research has advanced considerably over the past 25 years (Schmidt, 2006), but to date there is "no comprehensive set of biomarkers of carcinogen exposure or biological effects as a predictive measure of the total carcinogenicity related to exposure to tobacco or tobacco smoke" (Hatsukami *et al.*, 2006). The Institute of Medicine report *Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction,* cited the need for biomarker development in their principal research recommendations: "Although candidate disease-specific surrogate markers are currently available, further validation of these markers is needed. In addition, other biomarkers that accurately reflect mechanisms of disease must be developed to serve as intermediate indicators of disease and disease risk." (Institute of Medicine, 2001). Another expert committee, that assembled to identify key research needs related to tobacco harm reduction, also included among its recommendations the need to identify and validate biomarkers that are predictive of later disease development (Hatsukami *et al.*, 2002). Many biomarkers are currently used in research to study biologic effects of tobacco products or potential reductions in exposure from modified products. Table 5.21 lists a panel of biomarkers that have been recommended as the most promising for use in research on potential reduced exposure products. However, these biomarkers are not necessarily ready for use in a regulatory setting as they require better characterization of their relation to health risks and disease.

A candidate biomarker must go through a process of validation that establishes the qualitative and quantitative relationship of the biomarker to a specific exposure (i.e. a chemical in tobacco smoke) and to a selected end-point (i.e. cancer) (International Programme on Chemical Safety, 1993). There are several issues to consider in evaluating a candidate biomarker including: understanding of the role of the biomarker along a disease pathway, amount of supportive dose-response data (e.g. quantitative data correlating levels of the biomarker with smoking status and with disease endpoints), specificity (is it specific for exposure to tobacco toxicants?), sensitivity (are available tests sufficiently sensitive to detect quantities within a range encountered in the population and to detect meaningful changes), and reproducibility (e.g. intra-subject reliability) (Institute of Medicine, 2001). Supportive data for a biomarker's association with tobacco use should ideally include differences between tobacco users and non-users, a decrease with cessation of tobacco use, a dose-response relationship with quantity or frequency of use, and a decrease with reduced smoking (Hatsukami *et al.,* 2006). Additionally, identification of multiple biomarkers along a continuum from exposure to early disease effects can provide a more robust profile of the relationship between exposure and disease risk.

Biomarkers of internal exposure:

Biomarkers of internal exposure can potentially provide a more

General Tobacco Exposure
Nicotine/Cotinine
Carbon Monoxide
Cancer
NNAL
NNAL Glucs
3-Aminobiphenyl
4-Aminobiphenyl
Sister chromatid exchange
Nonmalignant Lung Disease
Macrophages
Cardiovascular Disease
Flow-mediated dilation
Circulating endothelial precursor cells
Fibrinogen
Homocysteine
White blood cell count
C-reactive protein
sICAM1
Glucose-clamping studies

Adapted from Hatsukami *et. al.* (2006)

*Held in February, 2004, and sponsored by the National Cancer Institute, the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, and the Centers for Disease Control and Prevention.

Table 5.21 Panel of Biomarkers: Recommended by 2004 Conference* on Methods and Biomarkers to Assess Potential Reduced Exposure Tobacco Products (PREPS)

accurate estimate of actual exposure received by the smoker than can be inferred from machine-based cigarette ratings or number of cigarettes smoked. For example, it was found that over an approximately 10-fold range in FTC cigarette ratings there was little or no significant difference in blood nicotine levels in several studies, demonstrating that FTC ratings do not reflect actual uptake of nicotine by the smoker (Benowitz, 1996b).

Nicotine metabolites have been widely used as biomarkers of general exposure to tobacco products, including exposure to smokeless tobacco and to environmental tobacco smoke (ETS [referred to in this volume as secondhand smoke (SHS)]) among nonsmokers (Benowitz *et al.*, 1994; Benowitz, 1999). Cotinine is the most widely used metabolite, as it has a relatively long elimination half-life of 16 hours (compared to only two hours for nicotine) and can be easily measured in urine, serum, or saliva. Nicotine has also been measured in hair and toenails as a means of assessing exposure to SHS in large scale epidemiologic studies, although the reliability of these measures may be influenced by hair treatment, and other factors, and requires further evaluation (Al-Delaimy, 2002; Al Delaimy *et al.*, 2002). Nicotine and its metabolites also make effective biomarkers because they are highly specific to tobacco exposure (unless the subject is using nicotine replacement therapy).

Carbon monoxide (CO) exposure has also been used as a biomarker for exposure to tobacco smoke. CO can be measured in exhaled air, as CO boost before and after cigarette smoking, and in blood as carboxyhemoglobin (Benowitz, 2003). While CO is not specific to tobacco, it can serve as a reliable short-term measure of smoking. The minor tobacco alkaloids anabasine and anatabine, which are specific to tobacco products and can be measured in urine, have also been used in studies for verifying smoking status (Jacob *et al.*, 2002).

Chemically-specific biomarkers can be used to assess exposure to particular toxins and carcinogens in tobacco and smoke, which may be valuable for evaluating the impact of product performance standards targeting specific constituents. Among the

chemical biomarkers, NNAL and its glucuronides (NNAL-Glucs), which are metabolites of NNK, are particularly useful because they are specific for exposure to tobacco products (as NNK is a tobacco-specific carcinogen) (Hecht, 2002). NNAL and NNAL-Glucs are measured in urine and have been used to quantify levels of NNK uptake in smokers and smokeless tobacco users, and to assess changes following cessation or product switching (Hecht *et al.*, 2002; Hatsukami *et al.,* 2004; Joseph *et al.*, 2005; Lemmonds *et al.*, 2005).

Biomarkers of potential harm:

DNA adducts potentially provide a direct measure of tobacco-induced DNA damage. Adducts are formed when chemical carcinogens bind to DNA, which can alter the structure of the DNA and is believed to be an important step in the pathway to cancer. Protein adducts have also been used to determine levels of carcinogen exposure and activity, since most carcinogen metabolites that react with DNA will also react with proteins, such as hemoglobin, and they are more readily measured than DNA adducts (Ogawa *et al.*, 2006). Hemoglobin (Hb) adducts of aromatic amines, particularly 3- and 4-aminobiphenyl, have shown promise for use in studies of tobacco-related carcinogen exposure. They have been shown to be higher in smokers than nonsmokers (Bryant *et al.*, 1987; Phillips, 2002), and have also been used to measure exposure to carcinogens in secondhand smoke (Hammond *et al.*, 1993). Aromatic amines are not specific to cigarette smoke exposure, however, and can also be associated with occupational and other chemical exposures.

Among the complex biomarkers of DNA damage and potential harm, urine mutagenicity and sister chromatid exchanges are the most promising as indicators of potential cancer effects. Both of these measures have been found to be higher in smokers than nonsmokers and to decrease on cessation (Vijayalaxmi & Evans, 1982; De Marini, 2004). However, the measured effects may be caused by diet or other factors, as well as cigarette smoke, and these differences may reflect other risk behaviour patterns associated with smoking. Development of complex measures that assess the combined effects of tobacco toxins and carcinogens is important because chemically-specific biomarkers, while they may have greater specificity in relation to exposure, may be misleading as a measure of disease risk. A reduction in uptake of a single tobacco smoke constituent in smokers, such as NNAL, may not necessarily provide any meaningful reduction in risk. Consumers may interpret a claim of reduction in a single chemical exposure as indicating a health benefit. Thus, such measures should be put in the context of overall hazard from a complex product.

Biomarkers of potential harm have been used in the research context, such as in clinical studies of potential reduced exposure tobacco products (Breland *et al.*, 2006). However, at this point, none of these biomarkers have been recommended for widespread use in regulation because their relationship to risk and health outcomes has not been sufficiently characterized.

Surveillance

Comprehensive surveillance is essential to assess the impact of regulation on tobacco product use and effects across the population. However, this remains a challenge because capacity and infrastructure for surveillance is limited in many countries (Jha & Chaloupka, 2000). Thus, the extent of surveillance efforts and available infrastructure is likely to vary widely between countries. A comprehensive surveillance programme could potentially cover an enormous range of information. Broadly, surveillance efforts should address changes in the design and performance of the product itself, marketing activity, beliefs and attitudes around tobacco product use, tobacco use behaviours, including initiation and cessation, and health outcomes. Suggested construct areas for post-marketing surveillance are drawn from published recommendations and are listed in Table 5.22 (Institute of Medicine, 2001; Hatsukami *et al.*, 2005).

In addition to measuring potential changes in specific tobacco constituent exposures, it is important to track tobacco

product use and risk beliefs in relation to product regulations. Product modifications in response to regulation may impact tobacco use behaviour. Additionally, experience with "light" cigarettes has provided substantial evidence that smokers believe these products to be less harmful (Cohen, 1996a; Giovino *et al.*, 2000; Ashley *et al.*, 2001; Shiffman *et al.*, 2001). Establishment of regulatory performance standards or constituent upper limits, for example, may be misinterpreted as "safe" levels of exposure. While laboratory evaluation of product design and emissions can provide early warning of potential adverse effects, comprehensive post-marketing surveillance is essential to ensure that regulations are achieving their aims. Additionally, independent technical and research capacity and infrastructure are needed to track changes in tobacco products and users' behaviour.

Establishing laboratory research and testing capacity is a crucial step in supporting surveillance activities to inform evaluation of tobacco product regulation. In addition to tobacco product regulations, governments may have research capacity for studying other aspects of tobacco products. The objective of standardised product testing is to assess product performance and characterize the delivery of particular constituents known to be important for public health, such as carbon monoxide, nicotine, and nitrosamines. In contrast, the goals of research efforts are to understand better the nature of tobacco products, how they work, their effects, and how they might be modified to alter their effects. While testing operations adhere to standardised protocols, research endeavors aim for flexibility and development of new methods and measures for ongoing scientific discovery and analysis. The WHO Study Group on Tobacco Product Regulation has highlighted how both research and testing capacity are essential and must be coordinated (WHO Study Group on Tobacco Product Regulation, 2004). For example, as tobacco products change, new products are introduced, and new scientific methods become available; therefore, it may be necessary to develop new performance standards. Additionally, previous efforts to promote product modification to protect public health, through lowering measured tar and nicotine yields in cigarettes, were undermined by a lack of expertise on tobacco products and smoking behaviour in the public health community (Parascandola, 2005). Thus far, tobacco testing and measurement standards have been primarily driven by the interests of the tobacco industry; thus it is important that the public health community develop capacity and expertise in this area to ensure that product regulations serve the aims of public health (Bialous & Yach, 2001).

In 2005, WHO convened the first meeting of the Tobacco Laboratory Network (TobLabNet), which included more than 25 laboratories from 20 countries. The primary goal of the meeting was to establish a global network of government, university, and independent laboratories to strengthen national and regional capacity for the testing and research of the contents and emissions of tobacco products pursuant to Article 9 of the WHO FCTC. Future activities of the network may include training programmes and development of common measures and protocols (http://www.who.int/tobacco/global_interaction/tobreg/ laboratory/en/ index.html). More details about recommended equipment, personnel, and resources for operating a tobacco product testing laboratory are provided by TobReg (2004). There is a substantial need for support and development of laboratory capacity independent of the tobacco industry in countries around the world with the purpose of achieving public health goals.

Summary

Articles 9 and 10 of the WHO FCTC call for ratifying nations to adopt policies for the regulation and disclosure of tobacco product contents and emissions. This chapter focuses on a review of the methods and measures for evaluating policies that are intended to regulate tobacco products. There are currently five main types: 1) regulations that require disclosure of product information; 2) regulations intended to reduce product toxicity and

Tobacco Product Design and Performance	Product contents, design features (filter, cigarette body), emissions of constituents that modify toxicity and addiction, additives, ignition propensity.
Marketing Activity	Product packaging and labelling, advertising content, promotional materials.
Beliefs and Attitudes	Product awareness, understanding of product design and regulation, risk perception, sensory responses.
Tobacco Use Behaviours	History, current use, brand use, quit attempts/history, addiction/dependence, readiness and intentions to quit, demographics, smoking topography.
Health Outcomes	Biomarkers of toxin exposures, biomarkers of early biological effects, tobacco-related disease incidence.
Other Outcomes	Fires caused by cigarettes.

Table 5.22 Surveillance Construct Categories

harm; 3) regulations intended to reduce the addictiveness and/or attractiveness of tobacco products; 4) regulations intended to prevent fires caused by cigarettes; and 5) bans (or removal of bans) on product categories. The selection of specific constructs and methods for evaluation will vary depending on the goals of the specific policy. However, as a general framework, it is likely that the impact of tobacco product regulations on intended health outcomes will be moderated by changes in product design, performance, marketing, product-related beliefs and attitudes, and tobacco use behaviour, which in turn are expected to influence exposures to tobacco constituents and emissions. Thus, evaluations should not be limited to assessing compliance within the intended effects of a regulation, but should also consider unintended effects of responses, such as tobacco industry innovation, that may interfere with the impact of the regulation.

There is a need for a centralized database that would, at a minimum, characterize different product regulations so that the effects of different policies can be compared. Additionally, as a condition permitting tobacco product sales, governments should require (if they do not already) tobacco product manufacturers to regularly disclose information about their products at the finest level of brand subcategory, including sales and marketing data, product content, and design features. This is needed to inform the development, implementation, and evaluation of effective regulations. Additionally, ongoing surveillance is required to assess the impact of tobacco product regulation on the tobacco product market and on the population, as well as to detect industry responses and other unanticipated consequences of regulation. The challenges of measurement associated with evaluating the effects of tobacco product regulations should not be underestimated. For example, many governments have enacted maximum smoke emissions standards (i.e. tar, nicotine, and carbon monoxide) based on standardised machine testing protocols for the purpose of reducing exposure to the constituents in tobacco products and resultant harm. However, based on the evidence reviewed in this Handbook, it is not recommended that yields from standard machine testing protocols, such as the ISO cigarette testing method (ISO Standard 3308:2000 (4th edition)), be used to assess or predict human

exposure. Emission yields derived from these protocols are not valid measures of actual human exposure. In order to evaluate the effectiveness of product regulations aimed at reducing harm, measures of human use and exposure are essential. There is an urgent need to identify valid methods and measures for assessing human exposure and harm that have practical utility for evaluating tobacco product regulations.

5.4 Measures to assess the effectiveness of restrictions on tobacco marketing communications

Introduction

The WHO FCTC proposes a comprehensive ban on tobacco advertising, promotion, and sponsorship, in recognition that it would reduce consumption of tobacco products (Figure 5.12). This section will explain how to go about measuring the effectiveness of restrictions on tobacco marketing communications, such as advertising bans or limitations on the use of specific media. First, terms are defined and explanations given on how promotional activity fits into the wider marketing strategy of tobacco corporations. The importance of restrictions on tobacco promotion is discussed, as well as the need to measure their effectiveness. Different ways of measuring effectiveness are looked at, with an argument that consumer surveys are one of the most useful. Finally, specific measures that can be used are offered.

Defining terms: tobacco promotion and marketing

Tobacco promotion covers all the communication efforts tobacco corporations use to encourage consumption of their products. These include mass media advertising (e.g. television, posters, and in the press), sponsorship of sporting and cultural events, point-of-sale promotion, merchandising and give-aways, and public relations. Table 5.23 provides an illustrative list.

The communication efforts, or more accurately *marketing communications*, outlined in Table 5.23, aim to encourage consumption of tobacco products by relaying a variety of messages to customers. As well as communicating basic product information and reminding the world about its product, marketing communications are used to reassure current customers that they have made the right decision, encourage new customers to try their product, and steer customers away from competitors. In essence the goal is to tell the customer or potential customer how the offering fulfils their needs.

A well-established business literature about the value of *integrated marketing communications* (IMC) (Schultz & Kitchen, 2000) argues for combining mass media and other *marketing communications* in a marketing communications mix. IMC holds that all company communications with their customers, through whatever channels, should be coordinated and coherent to articulate a completely unified message. In this way, the whole can become greater than the sum of the parts. For example, this comment from a tobacco industry advertising agent shows how merchandising, packaging, and advertising are pressed into joint service:

> "*What I would add is that there is a definite sub-culture among younger roll-your-own smokers, and I believe their desire to display their exclusivity could be supported by provision of unusually designed "badges" such as (transparent?) Raw lighters and rolling machines. This will enable them to differentiate themselves from uncool, older GV [Golden Virginia] smokers, who I suspect would not be particularly motivated to buy the product by either the advertising or the packaging"* (Collet Dickenson Pierce, 1999).

For many fast moving consumer goods (that have a quick turnover and relatively low cost), the ultimate aim of integrated marketing communications is to build evocative brands; something the tobacco multinationals do well, and is crucial for their financial success. Brands and their carefully crafted imagery are the principal means of meeting the psychosocial needs of one of their most important markets: young people. Ultimately, "*if a brand of cigarettes does not convey much in the way of image values, there may well be little reason for a young smoker to persist with or adopt the brand*" (Rothmans Marketing

1. Parties recognize that a comprehensive ban on advertising, promotion and sponsorship would reduce the consumption of tobacco products.
2. Each Party shall, in accordance with its constitution or constitutional principles, undertake a comprehensive ban of all tobacco advertising, promotion and sponsorship. This shall include, subject to the legal environment and technical means available to that Party, a comprehensive ban on cross-border advertising, promotion and sponsorship originating from its territory. In this respect, within the period of five years after entry into force of this Convention for that Party, each Party shall undertake appropriate legislative, executive, administrative and/or other measures and report accordingly in conformity with Article 21.
3. A Party that is not in a position to undertake a comprehensive ban due to its constitution or constitutional principles shall apply restrictions on all tobacco advertising, promotion and sponsorship. This shall include, subject to the legal environment and technical means available to that Party, restrictions or a comprehensive ban on advertising, promotion and sponsorship originating from its territory with cross-border effects. In this respect, each Party shall undertake appropriate legislative, executive, administrative and/or other measures and report accordingly in conformity with Article 21.
4. As a minimum, and in accordance with its constitution or constitutional principles, each Party shall:
   (a) prohibit all forms of tobacco advertising, promotion and sponsorship that promote a tobacco product by any means that are false, misleading or deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions;
   (b) require that health or other appropriate warnings or messages accompany all tobacco advertising and, as appropriate, promotion and sponsorship;
   (c) restrict the use of direct or indirect incentives that encourage the purchase of tobacco products by the public;
   (d) require, if it does not have a comprehensive ban, the disclosure to relevant governmental authorities of expenditures by the tobacco industry on advertising, promotion and sponsorship not yet prohibited. Those authorities may decide to make those figures available, subject to national law, to the public and to the Conference of the Parties, pursuant to Article 21;
   (e) undertake a comprehensive ban or, in the case of a Party that is not in a position to undertake a comprehensive ban due to its constitution or constitutional principles, restrict tobacco advertising, promotion and sponsorship on radio, television, print media and, as appropriate, other media, such as the internet, within a period of five years; and
   (f) prohibit, or in the case of a Party that is not in a position to prohibit due to its constitution or constitutional principles restrict, tobacco sponsorship of international events, activities and/or participants therein.
5. Parties are encouraged to implement measures beyond the obligations set out in paragraph 4.
6. Parties shall cooperate in the development of technologies and other means necessary to facilitate the elimination of cross-border advertising.
7. Parties which have a ban on certain forms of tobacco advertising, promotion and sponsorship have the sovereign right to ban those forms of cross-border tobacco advertising, promotion and sponsorship entering their territory and to impose equal penalties as those applicable to domestic advertising, promotion and sponsorship originating from their territory in accordance with their national law. This paragraph does not endorse or approve of any particular penalty.
8. Parties shall consider the elaboration of a protocol setting out appropriate measures that require international collaboration for a comprehensive ban on cross-border advertising, promotion and sponsorship.

WHO (2003)

Figure 5.12 WHO FCTC Article 13: *Tobacco Advertising, Promotion and Sponsorship*

Advertising
Broadcast media (TV, radio, cinema)
Outdoor (billboards, posters outside stores)
Press

Sponsorship of Sports and the Arts

Point-of-Sale
Promotional material in shops (branded gantries, clocks, signage, staff clothing)

Coupon Schemes
Coupons included in packs of cigarettes that can be exchanged for free gifts

Merchandising
Low cost items (pens, lighters or t-shirts), competitions, free cigarettes

Special Price Offers
Short-term low price offers advertised in-store, on pack flashes, or in other media

Promotional Mail
Marketing communications sent straight to customers

Brand Stretching
Non-tobacco products with tobacco branding (Marlboro Classic clothes)

Pack designs to communicate brand image and to add value

Internet sites
Websites promoting tobacco companies, cigarette brands, or smoking

Product Placement
Paid for placement of cigarette brands in films or television

Table 5.23 Examples of Tobacco Marketing Communications

Services, 1998). The challenge therefore is to "*cement the brand into the repertoire of the experimental smoker*" (Collet Dickenson Pierce, 1996).

It has been found that younger smokers give more weight to the imagery of cigarettes, and pay more attention and are receptive to fashionable brands and the latest designs (Hastings & MacFadyen, 2000). Well-known brands, most notably Marlboro lights, exploit these emotional needs and insecurities: "*the success of Marlboro Lights derives from its being…the aspirational lifestyle brand… "cool"…the Diet Coke of cigarettes*" (The Leading Edge Consultancy, 1997).

The power of brand imagery is not only used on the young. In the low tar sector, branding, names, and liveries are used to create reassuring images and associations. For example, a low tar product "*is supported by the brand's imagery," which has a "high association with 'health conscious people*'" (Marketing Trends, 1995). Also, the tobacco industry has used images of happiness, physical well-being, harmony with nature, and a self-image of intelligence to appeal to the older, "concerned" smokers to discourage them from quitting (Pollay, 2000; Pollay & Dewhirst, 2002).

Evocative branding, created through research to complement consumers' lifestyles and aspirations, is spread by integrated marketing communications. This communication effort dovetails with the company's wider marketing effort, encompassing product design, pricing, and distribution, to ensure optimal consumer satisfaction.

The product's marketing function is reinforced by its prominence in the smoker's life: "*Smokers buy cigarettes frequently. They carry their brand around with them and see other brands constantly. The product is a prime means of communicating a change*" (Collet Dickenson Pierce, 1998).

New product development ideas ensure that the needs of consumers are met and that appropriate pharmacological and aspirational benefits are offered. The new smoker is assisted on the passage from experimenter to regular smoker by lower pH levels in cigarettes, which lowers the rate of absorption of nicotine, thus minimising the initial side-effects of smoking, such as dizziness and nausea (Claude, 1973). Tobacco marketers have also developed "product line extensions" specifically in response to increasing health propaganda. For virtually every brand there is now a "light" or "low" alternative, providing the worried smoker with an excuse or rationalisation to continue smoking. Other development ideas include an Espresso cigarette to fit the new "café culture" and to provide "quick hit (caffeine/nicotine) with young, streetwise imagery" (e.g. "a lad's cigarette, complete with scantily clad women on the cigarette paper!"), and "nationalistic (but not jingoistic)" Scottish and Welsh cigarettes to exploit devolution (Hastings & MacFadyen, 2000). These ideas never reached the street, but they do illustrate how the product is manipulated to create synergy with the overall marketing effort.

Pricing strategies are also important to tobacco companies, and the relationship between quality, brand image, and price is particularly so, as it feeds into fundamental decisions about segmentation and targeting. Thus, for the starter segment, premium pricing is appropriate. While adolescents tend to be more price sensitive than adults, they attribute a greater premium to the image attached to the more expensive product, if they are visible and socially important. Therefore the pricing strategy should clearly demonstrate the high quality and style of the brand, if the product is to meet the adolescent's needs for image and social status (DiFranza *et al.*, 1991; DiFranza, 1995; Barnard & Forsyth, 1996; Pollay *et al.*, 1996).

For established smokers, their addiction and maturity makes the price-quality relationship less of an issue, making them more willing to trade down. In response, the industry runs coupon schemes and sales promotions to reduce the perceived price of smoking. These types of pricing strategies tie the established smoker to one particular brand and reward them for their loyalty.

For the tobacco industry, the distribution system helps build the brand personality and target the specific need of each segment. Despite bans on the sale of cigarettes to minors, distribution tactics still play a big role in targeting them. Wide distribution ensures cigarettes become omnipresent and a cultural norm, encouraging adolescents to overestimate the extent, and underestimate the social disapproval, of smoking (Davis, 1991; Wakefield *et al.*, 1992; Evans *et al.,* 1995). More prosaically, marketers can place their products in those outlets where it is easier for adolescents to buy cigarettes and many of them do so successfully. In the UK, outlets such as newsagents, tobacconists, and sweetshops are the most popular source for sales to young smokers (Boreham & Shaw, 2001; Bates *et al.*, 2005), making them a good option for under-age distribution.

For the established smoker, wide distribution also helps create an environment of normalcy and reassurance. Furthermore, the distribution network is so complete that the smoker can rest assured that cigarettes will always be readily available.

Thus, the industry's use of integrated marketing communications is nested in their wider marketing effort involving a consumer oriented strategy to get *"the right product, at the right time, in the right place, with the right price*" (Cannon, 1992).

The issues of product design and pricing, and how these can be

measured, are discussed in Sections 5.3 and 5.1, respectively. This section is concerned with examining marketing communications; the evidence base that shows that these strategies do influence smoker's behaviour and that they need to be restricted.

Why restrictions on tobacco marketing communications matter

To help understand the potential effect of removing or restricting tobacco marketing communications, it is helpful to first look at a selection of studies that have examined the influence exerted by tobacco marketing, and the approaches and measures that have been used in these studies. They can be helpful, not only in guiding expectations about attributes that will change as a result of restrictions, but also in identifying which measures are important to collect.

Modelling aggregate demand:

One of the first and most influential studies into the effects of tobacco promotion on consumption was conducted in the UK (McGuinness & Cowling, 1975). It modelled the aggregate demand for cigarettes in terms of price, income, and advertising. The advertising measure was an estimate of the number of messages received by a consumer rather than expenditure. Their findings suggested that advertising does have a significant effect on cigarette sales, but that publicity of adverse health effects of smoking had reduced the sales effect of cigarette advertising.

Evaluation of advertising bans:

Evidence from studies evaluating the effects of advertising bans also show that marketing communications have a significant effect on consumption. The Smee Report, which analysed Norway's 1975 Tobacco Act, concluded that the Act decreased smoking demand between 9% and 16% (Economics and Operational Research Division of the Department of Health in England, Smee *et al.*, 1992 - Economics and Operational'... England). Similarly a study of the effects of the 1971 Finland Tobacco Act, which analysed data from 1960 to 1987, concluded that the advertising ban produced a long-term reduction of 6.7% in cigarette smoking (Pekurinen, 1989). Measures of per capita annual consumption of cigarettes and tobacco were analysed by extent of advertising bans across 22 countries (Saffer & Chaloupka, 2000). Minimal effect was found from limited bans in reducing tobacco use, but clear effect from comprehensive bans. (See the following section on "Advertising bans of specific media" for definitions of the types of bans.)

Evaluation of individual campaigns:

Evaluation of individual campaigns reveals how the tobacco industry has targeted specific groups. For example, an evaluation of a Camel cigarette campaign in the early 1990s revealed that in a short period of time, it had made a huge impact upon children's smoking behaviour (DiFranza *et al.*, 1991). The campaign featured a cartoon drawn Camel, known as Joe the Camel, which was suspected of having particular appeal to children. The study asked about brand preference and compared it with data from seven surveys conducted prior to the launch of the Camel campaign. In the three years following the start of the campaign there was an increase from 0.5% to 32.8% in the proportion of young smokers (aged up to 18) who named Camel as their preferred brand. The study measured awareness of the campaign and identification of product type and brand name by showing an advert masking all clues (except Old Joe) to the product and brand being advertised. The research found that children were more aware of the campaign and more able to identify the product type and brand name from the logo than adults. A campaign "appeal score" was compiled by asking subjects to rate six unmasked Old Joe adverts across four items: cool, stupid, interesting, or boring. They were asked if they thought Old Joe was "cool" and if they would like to be "friends" with him. Positive responses to each item were scored 1 and negative responses coded 0 and the appeal score was the arithmetic sum of these. Children were found to be more

likely than adults to find the campaign appealing.

Brand awareness and appreciation:

Campaigns are also linked with increases in brand awareness and appreciation. Qualitative work was conducted which found that children, as young as six years old, were aware of cigarette advertising, and that young primary school children had learned the brand imagery or personality of leading cigarette brands from advertisements (Aitken *et al.*, 1985). A survey in England showed that 17% of 9-10 year olds and 23% of 12-13 year olds were able to name a favourite cigarette advertisement (Charlton, 1986). The brands named most frequently were also those most heavily advertised in the area at the time. In addition, it was found that the children who named favourite advertisements were also more likely to agree with some positive statements about smoking and the image of smokers. It concluded that children receive positive messages about smoking behaviour from advertising, which may reinforce their decision to start smoking during experimentation. Thus, if tobacco advertising is banned, the expectation is that these positive messages will lessen or be eliminated.

Brand choice:

Studies have also examined brand choice in relation to tobacco advertising. Young smokers tend to be particularly attracted to the most heavily advertised products, and it is these brands that dominate under-age sales. For example, the three most heavily promoted brands in the USA in 1993 (Camel, Marlboro, and Newport), were the three most likely to be purchased by adolescents (Centers for Disease Control and Prevention, 1994c). Similar patterns of preference for heavily advertised brands have also been observed amongst UK adolescents (Barton, 1998), and prolonged advertising makes brands seem popular (Sutherland & Galloway, 1981).

Although the majority of studies have examined the impact of mass media advertising on smoking, many other forms of marketing communication have also been studied (see Table 5.24). It is important to keep in mind the discussion about integrated marketing communications at this point, as none of these communication efforts are intended to work in isolation. Indeed the final study listed in Table 5.24 underlines this point by demonstrating a *cumulative* impact: the more forms of marketing communications that young people are aware of, the more likely they are to smoke.

Longitudinal designs:

The research discussed thus far has provided convincing evidence that there is a relationship between tobacco marketing communications and smoking behaviour. However, it has not established cause and effect; longitudinal designs are needed to do this. A longitudinal study was undertaken to measure the predisposing effects of cigarette advertising on children's intentions to smoke when they were older (Aitken *et al.,* 1991). Two interviews were conducted among children aged 11-14 years: those who expressed a stronger intention to smoke during the second interview rather than the first, were more likely to have liked cigarette advertising at baseline. This demonstrates that nonsmokers, who felt that they may smoke when they were older, were paying more attention to cigarette advertising than other nonsmokers.

An important meta-analysis of longitudinal surveys has recently been published by the Cochrane Library (Lovato *et al.*, 2003). The authors asked the question: "*is prior exposure to tobacco industry advertising and promotion associated with future smoking among adolescents?*". They analysed the outcome of nine longitudinal studies, including the study mentioned above. All nine studies showed "a positive, consistent, and specific relationship" between exposure to tobacco advertising and influence upon adolescents to smoke cigarettes. The authors concluded:

> *"Longitudinal studies suggest that exposure to tobacco advertising and promotion is associated with the likelihood that adolescents will start to smoke. Based on the strength of this association, the consistency of findings across numerous observa-*

Sponsorship

- Exposure to a cigarette sponsored sports advertisement reinforced existing smoking behaviour, and for non-smokers created favourable attitudes towards smoking, increased awareness, and liking of brands (Hoek *et al*., 1993)
- Children show a higher awareness of the sponsoring brand, and link the exposure to brand recall and understanding of brand imagery (Ledwith, 1984; Aitken *et al.,* 1986; Piepe *et al*., 1986)
- Children's preference for motor racing is a significant independent variable in move to regular smoking (Charlton *et al*., 1997)
- The statement "smoking can't be all that dangerous, or the Government would ban sports sponsorship" was put to over 4000 11-16 yr olds; substantially more smokers than nonsmokers agreed with it (Bates, 1999)

Merchandising

- Items such as branded lighters, t-shirts, baseball caps, and badges frequently reach adolescents at the point-of-sale, special events, or through competitions (Coeytaux *et al.,* 1995; Gilpin *et al*., 1997; Pierce *et al*., 1999)
- There is a significant relationship between experience of tobacco promotions and susceptibility to tobacco use (Altman *et al*., 1996; Gilpin *et al*., 1997; Feighery *et al*., 1998)
- There is a relationship between the numbers of promotional items owned and a higher likelihood of smoking (Sargent *et al*., 2000)
- There are relationships between smoking initiation rates and levels of promotional expenditure, and owning/using tobacco promotional items and the onset of smoking (Bauer & Johnson, 1999; Redmond, 1999)

Brand-Stretching

- For example, the endorsement of holidays, cafés and music; items that are then sold rather than given away (Centre for Tobacco Control Research, 2001)
- Initial research focussed mainly on advertising for such products, and shows that this is consistently seen as advertising for the sponsoring tobacco brand rather than the product (Aitken *et al*., 1985; Centre for Tobacco Control Research, 2001)
- The awareness of brand stretching by 15 year olds is independently associated with being a smoker (MacFadyen *et al*., 2001)

Packaging

- Tobacco packaging both reinforces brand imagery and reduces the impact of health warnings (Beede & Lawson, 1992; Carr-Greg & Gray, 1993; Goldberg *et al*., 1995; Rootman & Flay, 1995)
- When fewer brand image cues were on the packaging, adolescents were able to recall more accurately non-image health information (Beede & Lawson, 1992)
- Plain packaging limits the ease with which consumers associate particular images with cigarette brands and significantly influences smoking behaviour (Goldberg *et al*., 1995)

Point-of-Sale (POS)

- Cigarette packets were displayed in such a way at the POS as to act like advertising (DiFranza *et al.,* 1999)
- Young adolescents who reported seeing tobacco advertising in stores were 38% more likely to experiment with smoking, and the advertising was found to enhance brand imagery (Schooler *et al.,* 1996; Donovan *et al*., 2002)
- The more youth-orientated ads were displayed outside shops, the more often children tried to buy cigarettes (Voorhees *et al*., 1998)
- There are greater levels of POS advertising in areas where there is likely to be a high prevalence of smoking (e.g. low-income / ethnic minority areas); young people are unduly exposed to them (Woodruff *et al*., 1995; Ruel *et al*., 2001; Laws *et al*., 2002)

Table 5.24 The Influence of Marketing Communications on Smoking Behaviour

Product Placement

- The paid for placement of cigarette products in films and on TV is a controversial, but documented, marketing communications tactic. Strong evidence links this with adolescent smoking (Hart, 1996; Chapman & Davis 1997; Dalton *et al.*, 2003)

Loyalty Schemes

- There is significantly greater participation in low-income areas, and coupons may offset the effect of price increases (Centre for Social Marketing, 1995)
- Loyalty schemes involvement among 15 year olds is independently associated with smoking (MacFadyen *et al.,* 2001)

Free Samples

- A systematic search of tobacco industry documents confirms free samples as a popular strategy (Sepe *et al.*, 2002)
- Receipt of free samples by young people independently associated with susceptibility to smoke (Altman *et al.*, 1996)

Internet

- Tobacco manufacturers have their own websites and sponsor further sites unrelated to tobacco. Also pro-tobacco sites (not related to industry) include chat rooms/message boards and celebrities/attractive role models smoking, which may appeal to the young (Center for Media Education, 1997; Center for Media Education, 1998; Hong & Cody, 2002)

Marketing Communications

- Young people are aware of *all* forms of tobacco marketing communications; over half of all smokers had participated in some form of promotion; and the greater the number of tobacco marketing techniques a young person was aware of, the more likely they were to be a smoker (MacFadyen *et al.*, 2001)

Table 5.24 The Influence of Marketing Communications on Smoking Behaviour

tional studies, temporality of exposure and smoking behaviours observed, as well as the theoretical plausibility regarding the impact of advertising, we conclude that tobacco advertising and promotion increases the likelihood that adolescents will start to smoke. From a policy perspective, attempts to eliminate tobacco advertising and promotion should be supported."

A useful codicil could be added to the authors' final sentence: that there is also a need to devise robust methodologies to monitor the effectiveness of any such prohibitions. To a large extent, the studies mentioned above have concentrated on measuring the influence of advertising. When measuring the effects and effectiveness of tobacco marketing restrictions/bans it is important to consider all potential forms of remaining tobacco marketing, and thereby monitor whether or not the tobacco industry diverts their marketing activities to less restricted media.

Alternative methodologies

In discussing the evidence base, it is apparent that various approaches and measures have been used to examine the effects of tobacco promotion. These same approaches are relevant and provide guidance as studies are designed to assess the effects and effectiveness of restrictions on tobacco marketing communications. Below, the two main approaches (econometric studies and consumer surveys) for examining the effects and effectiveness of tobacco marketing

restrictions are discussed. In addition, complementary approaches are addressed, including marketing surveillance and internal document analysis that can help to contextualise, interpret, and support results that emerge from consumer surveys and econometric analysis.

Econometric studies:

One approach is to use econometric[1] studies that model changes in tobacco consumption with fluctuations in tobacco advertising expenditures. There are two main types of econometric studies: comparative studies of countries with different levels of controls on advertising (cross-country studies); and studies which model the effect of year-to-year fluctuations in advertising expenditure on consumption within one particular country (time-series studies).

Prior econometric studies of tobacco consumption have used one of three alternative empirical measures of advertising: national aggregate expenditure data, cross-sectional measures of advertising, and advertising bans (Saffer & Chaloupka, 2000).

National aggregate expenditure data:

Annual national advertising expenditures are the yearly total of all cigarette advertising expenditures, for all advertisers, in all media, for all geographic market areas. However, the high level of aggregation of such data results in it having very little variation, which leaves little to correlate with consumption. It is therefore unlikely that any effect of advertising will be found from use of this type of data.

Cross-sectional data:

The types of cross-sectional data can vary, but would typically be local level (e.g. Metropolitan Statistical Area (MSA)) and for periods of less than a year. This type of data can have greater variation than national level data, as the cost of advertising, the mix of target markets, and relative size varies across local areas. Monthly or quarterly local level data would include a relatively larger variation in advertising levels and in consumption data, and be more likely to find a positive relationship between advertising and consumption.

However, cross-sectional studies are rare as the data are expensive and difficult to assemble. A report on 21 prior empirical studies, three of which were cross-sectional, found that in each of the three cross-sectional studies, a significant positive effect of advertising was observed (Saffer & Chaloupka, 2000).

Advertising bans in specific media:

Tobacco advertisers use a number of media, and while each has particular advantages and disadvantages, a partial advertising ban will likely result in tobacco advertisers substituting a banned media with a form of media that is not banned. A partial ban, therefore, will not necessarily imply a reduction in total expenditure on tobacco advertising. For example, in the USA advertising expenditure fell subsequent to the 1971 TV ban, but rose quickly thereafter. Three studies of advertising bans that used pooled international data were reported (Saffer & Chaloupka, 2000). Two of these studies showed no effect of a ban, while one showed that advertising bans had no effect on consumption in the period prior to 1973, but thereafter, cigarette advertising bans and warning labels had a significant negative effect on consumption. Studies that use advertising bans as the measure of advertising must therefore include bans which are sufficiently comprehensive to ensure that the industry cannot compensate for lost media by increasing advertising or other marketing expenditures. Changes in the number of countries having enacted more comprehensive tobacco advertising bans since the late 1980s provided the

[1]Application of mathematical and statistical techniques to economics in the study of problems, the analysis of data, and the development and testing of theories and models.

opportunity to re-examine the effects of advertising bans on tobacco consumption (Saffer & Chaloupka, 2000). Comparable economic and social data were available from 1960 for the 22 Organisation for Economic Cooperation and Development (OECD) countries studied. Four dependent variables were used in the regressions: two measures of per capita annual consumption of cigarettes, and two measures of per capita consumption of tobacco by weight. The data came from Health New Zealand and the United States Department of Agriculture. Advertising ban variables were created from data on television, radio, print, outdoor, point-of-purchase, and movie advertising, as well as sponsorship bans. These were converted into a set of three dummy variables: "Weak Ban" was set to equal one if there were zero, one, or two bans in effect; "Limited Ban" was set to equal one if there were three or four media banned; and "Comprehensive Ban" was set to equal one if there were five, six, or seven media banned. The analysis allowed assessment of the effect of limited and comprehensive bans, indicating minimal effect from limited bans in reducing tobacco use and clear effect on consumption from comprehensive bans (Saffer & Chaloupka, 2000).

Econometric studies of advertising and consumption are complicated and have produced mixed results. Part of the difficulty lies in the complexity of the procedure; models must account for a large number of other social, political, and economic factors, which may have a confounding effect on consumption patterns. Availability and completeness of data can also be problematic. Independent researchers, in the UK for example, have had to work within the limitations of incomplete advertising data released by the tobacco industry. The data provides coverage of broadcast media and the press, but omits billboard advertising and sponsorship. Studies in the USA, however, benefit from comprehensive data on advertising expenditure which is freely available to independent researchers, albeit in aggregated form. In the absence of suitable data for advertising, dummy variables can be used as proxies, like the dummy variables for strength of ban discussed above.

The inability of econometric studies to examine all the forms of marketing communication used by the tobacco industry, such as loyalty schemes or point-of-sale displays, was examined (Chapman, 1989). In addition, there are two further drawbacks with econometric studies: they only examine the effects of advertising on overall sales, ignoring other important influences on smoking-related cognition and beliefs; and they usually only provide aggregated, population level data; in most cases they are not able to examine effects on sub-groups (e.g. young people, women, or those on low income), some of whom may be particularly vulnerable.

Consumer surveys:

Another approach to examine the effects of tobacco promotion is through consumer surveys, which can overcome many of the problems associated with econometric studies. Consumer surveys can be appropriately timed to collect measures prior to the introduction of marketing restrictions and at a number of subsequent time points. At least one baseline measure is required prior to policy introduction, against which future changes can be gauged. The number and timing of post measures will depend on the timing of restrictions being posed and on the rate of change witnessed.

Consumer surveys allow social scientists to develop and test multiple hypotheses about tobacco marketing communications, the policies designed to restrict them, and how they may be working. In this way, specific sub-groups and a range of variables can be studied. Whereas econometric studies tend to rely on aggregate data, consumer surveys enable hypotheses about marketing communications to be tested at a more individual/disaggregated level, taking into account influences of individual characteristics.

This thinking can be built into a conceptual model, as with The International Tobacco Control Four Country Study (ITC) (Fong *et al.*, 2006a), where policies are characterised as potentially affecting individuals along a variety of psychosocial and behavioural variables, of which

there are two classes: policy specific variables and psychosocial mediators.

Policy specific variables are those that are proximal (conceptually closest), or most specifically related to the policy itself. For example, graphic warning labels should increase the prominence and noticeability of warnings, price should affect the perceived costs of cigarettes, and lifting of restrictions on alternative nicotine products should lead to increased awareness of their availability (Fong *et al.*, 2006a). Discrete behavioural changes may also occur as a result of the policy, such as smokers hesitating, or even abstaining from cigarettes because of the warning label.

Similar examples can be drawn in marketing communications. Restrictions on these should lead to reductions in awareness of the specific types of communication that have been restricted, such as billboards or press ads. Given the links found between tobacco advertising awareness and brand awareness and appreciation (Aitken *et al.*, 1985), restrictions on marketing may also reduce familiarity with tobacco brands.

Psychosocial mediators are those variables that are distal (conceptually distant) from the policy, and which are thought to be affected by multiple means, not just policies. Self-efficacy and intentions are amongst such variables. It is thought that policies will affect these general mediating variables indirectly, through their prior effect on the policy-specific, proximal variables (Fong *et al.*, 2006a).

The ITC conceptual model includes proximal and distal measures so as to construct a causal chain model. The route from policy specific variables to behaviour can be traced through these measurements. For example, withdrawal of tobacco marketing communications may first decrease awareness of tobacco marketing activity, which may then affect awareness and familiarity with brands, perceptions of smoking norms, overall attitudes, intentions about quitting (or intention to smoke among young people), and ultimately effect behaviour, such as quit attempts, quit success or, among young people, uptake of smoking (Figure 5.13). This model allows researchers to test how policies impact or fail to impact anticipated behaviour.

Three different studies have been undertaken to assess tobacco marketing restrictions: the ITC Four Country Survey (Thompson *et al.*, 2006), the Centre for Tobacco Control Research (CTCR) study (http://www.ctcr.stir.ac.uk), and the Global Youth Tobacco Survey (GYTS) (The Global Youth Tobacco Survey Collaborative Group, 2002).

The ITC project brings strengths to the consumer survey design. It is longitudinal, which enables disentangling cause and affect relationships. As a multinational study covering more than one jurisdiction, it allows for a quasi-experimental design: comparisons can be drawn between countries where specific policies are being introduced and others where they are not. It also is a telephone survey, which brings benefits in terms of sampling and ease of respondent access, but limits the complexity of the questions that can be asked because it is not possible to use show cards or any visual images.

The ITC project is conducted with adult smokers, and therefore examines effect amongst those already involved with tobacco products. A sample of over 2000 adult smokers is sought in each country at each wave of fieldwork (Thompson *et al.*, 2006).

As previously noted, consumer surveys have the advantage of enabling specific sub-groups to be studied, such as young people; some of whom will already be involved with tobacco products and some of whom will not. Consumer surveys, unlike econometric studies, enable the potential impact of tobacco marketing restrictions to be examined separately for young people. In particular, they enable the examination of how young people growing up in an environment surrounded by tobacco marketing compare with those growing up in an environment in which tobacco marketing is restricted.

The CTCR study and the GYTS focus on youth. The CTCR study is an ongoing, face-to-face,

in-home survey which, though logistically challenging, enables complex questioning procedures (particularly the use of visual aids displaying brand colours and design features). It uses cross-sectional surveys of 11-16 year olds across the UK. Surveys are conducted at approximately two yearly intervals to monitor changes in key measures (such as awareness of tobacco marketing, engagement with tobacco marketing, brand awareness and familiarity, perceived smoking prevalence, intentions to smoke, and smoking behaviour) at different time points prior and subsequent to the implementation of the UK ban on advertising and promotions (Tobacco Advertising and Promotions Act, 2002; http://www. opsi.gov.uk/acts/acts-2002/ukpga_20020036_en_1).

Approaches to measuring the marketing related measures are discussed below.

Two baseline surveys were conducted: two years prior to the ban and six months prior to the ban. These surveys provide data on young people's response to tobacco marketing prior to the regulations. Measures taken approximately 18 months post-ban gave an indication of short-term response following the initial phases of implementation of the tobacco marketing restrictions. Continuation at two year intervals will provide insights into the potential longer-term impact of these restrictions, and give an indication of the length of time before impacts may become apparent.

The GYTS is a school-based survey of 13-15 year olds which began in 1999 (The Global Youth Tobacco Survey Collaborative Group, 2002). It includes questions on prevalence of cigarette and other tobacco use, attitudes toward tobacco, access to tobacco products, exposure to secondhand smoke, school curricula on tobacco, media, advertising, and smoking cessation. The question focus, in relation to marketing restrictions, is on marketing penetration: awareness of media messages and receipt of tobacco branded items/gifts. Like the ITC Four Country study, the GYTS also uses multiple countries, a common methodology, and a core questionnaire, which has the potential to allow comparison across different levels of tobacco control (for details see Section 4.3)

Limitations:

The consumer survey approach has its limitations. It relies on gaining access to and cooperation from a representative sample of respondents, and on self-report measures which participants may under- or over-report.

The CTCR study is a national study conducted over a number of years. It monitors response to tobacco control policies as they change over time, providing data on reactions at different time points following staged implementation of the tobacco advertising and promotions ban. The lack of a comparison country or countries means that it cannot provide conclusive evidence concerning the impact of tobacco control policies. Nevertheless, it is a valuable study that can add to the understanding of the likely effect of marketing restrictions, particularly where consistencies and overlaps can be seen with the ITC Four Country study.

Enhancing benefits of consumer surveys:

The benefits of consumer surveys are enhanced when complementary methods are used to measure both marketing and policy inputs; a clear notion of what is happening out there will enhance the ability to measure its effectiveness. The policy and marketing arenas need to be systematically monitored in order to gauge the effect of developments. For example, Figure 5.13 demonstrates some of the responses that the tobacco industry may take. First, there is the issue of checking compliance, but equally, if not more important, is being aware of the innovative ways the industry may compensate for newly imposed restrictions.

Research tools, which aid work in the fields of surveillance, industry document analysis, and policy tracking, have been developed which enable the measurement of inputs. Monitoring these inputs also assists in contextualising and interpreting results from the consumer surveys and may help to clarify any unusual or unexpected survey results.

Multiple studies can also help to complement and reinforce results from individual surveys.

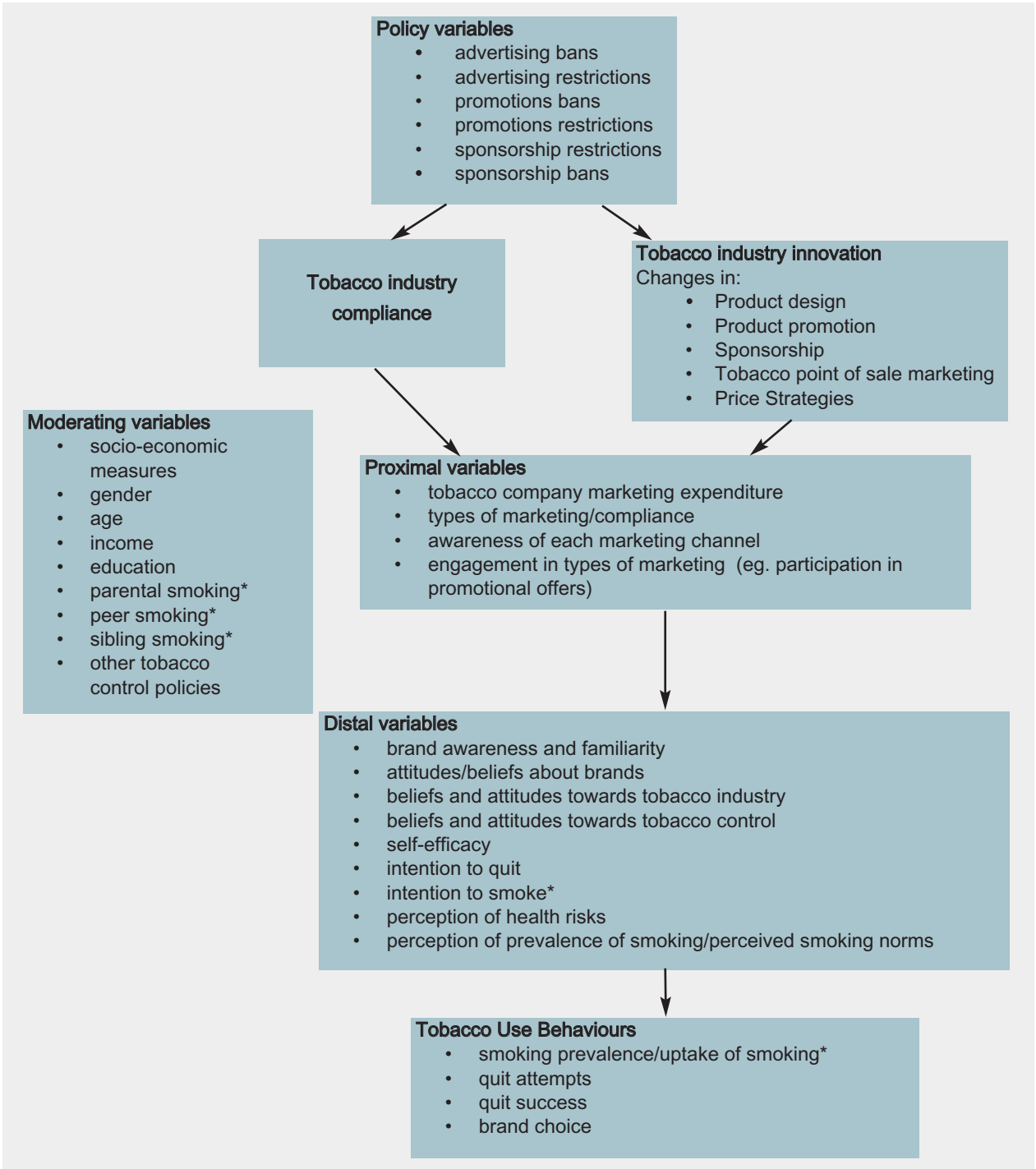


Figure 5.13 Conceptual framework for the evaluation of tobacco marketing restriction policies

*Appropriate measures for youth/adolescent studies

For example, while the CTCR study is an appropriate design for examining the responses of adolescents to the advertising and promotions ban, the lack of a comparison country limits the conclusions which can be drawn. The ITC Four Country survey monitors similar issues with adult smokers and does include comparison countries. Therefore, where findings are consistent between the two studies, the ITC Four Country study helps lend support for any findings from the CTCR study that are indicative of policy effects.

Marketing surveillance:

As part of the CTCR study, a series of marketing surveillance activities were undertaken to provide information about the marketing approaches being used by the tobacco industry both before and after the ban. This exercise was designed to capture the range and nature of activities, rather than quantify the amount of marketing activity. An observational protocol was developed to explore the tobacco industry's response to regulation in retail outlets (Devlin *et al.*, 2006); all other forms of tobacco marketing communications had been prohibited, but point-of-sale was still available. The protocol was designed to be generic to allow it to be adapted to cross-country comparisons, and to be executed longitudinally so long-term patterns could be uncovered. It was mainly comprised of closed questions requiring the trained observer to check the applicable box. Table 5.25 lists the types of measures recorded within the stores.

A small panel of 28 retailers were recruited to participate in this protocol, and a trained observer visited every two months. The panel consisted of a sample of different store types, but was not intended to be a representative sample. Rather, this study sought insight into the range of tobacco marketing at point-of-sale and how this might change over time and in response to new restrictions.

This observational protocol could be implemented with a much larger and representative sample of stores to enable comparison of data by different store types, area types, religions, and additional characteristics. Such an approach would enable comparison of data by different store types, different area types, regions, and so on.

Observations were conducted in all 53 stores that sold cigarettes within the study community (Feighery *et al.*, 2006). Two surveyors used a protocol for counting and categorising cigarette marketing materials and shelf space allocated to cigarettes in stores. Counts were made of the features, such as number of branded signs, merchandising fixtures, and functional items, along with amount of shelf space allocated to the three most popular cigarette brands among youth in the USA. These data were used alongside survey responses to assist with development of multiple measures of adolescents' exposure to retail cigarette marketing. Survey participants were shown photographs of the stores' exteriors, along with the names and addresses of 12 of the stores, and were asked to indicate the frequency of going to each one. Those who reported at least weekly visits to any of the specific stores in the photographs were classed as having frequent exposure to cigarette marketing.

This survey data was then combined with the observation to calculate a measure of "cigarette brand impressions per week." This was computed by multiplying the frequency of visits to the specific stores by the total number of marketing materials and product facings in each, and then summing all the individual store scores for each student.

Additional surveillance undertaken within the CTCR study includes regular audits of the press to identify any marketing or editorial coverage of tobacco products or issues. A selection of the most widely read newspapers and magazines are purchased over a one week period each six months and are content analysed for coverage of tobacco or smoking. A bi-monthly audit of the retail press is similarly undertaken to provide insight into the type of communication and messages being relayed from the tobacco industry to the retailers. A small panel of about 28 smokers also complete a form each month recording any tobacco marketing that they encounter, as well as recording their cigarette/tobacco purchases over a one week period. This gives an idea of the

Store Information
Characteristics of surrounding area (residential or commercial)
Presence or absence of tobacco ads on exterior of store
Presence or absence of minimum age of purchase signage
Size of outlet (number of cash registers)
Whether tobacco products or counter visible on entry to store
Cigarette/Tobacco Availability
Visibility, variety of brands, variety of pack sizes, presence of any promotions
Positioning of tobacco products
Which brands are most prominent
Availability, price, and price promotion of four particular brands
Advertising and Other Tobacco Marketing Practices
Presence or absence of advertising for four particular brands
Types of promotions observed in store and associated brands
Presence or absence of tobacco branded accessories
Use of Functional Objects at Point-of-Sale
Method of displaying tobacco products
Features of the cigarette display cabinet
Any noticeable changes in the cabinet
Presence or absence of tobacco branded fixtures or fittings in the store
Presence or absence of tobacco control signage

Table 5.25 Types of Measures Recorded within Retail Outlets to Monitor the Tobacco Industry's Response to Regulations

wide range of marketing activities, and can include, for example, promotions that occur in night-clubs, direct mail, free gifts, and special price offers.

Internet:

Given the increasing restrictions on other routes for tobacco marketing, the internet requires careful monitoring. A representative sample of websites with a dominant tobacco theme were researched (Hong & Cody, 2002). Three lists of search terms were generated: general smoking terms (category A), terms commonly associated with smoking (category B), and brand names of American tobacco corporations (category C) (see Table 5.26). All of the terms in category A, and all combined terms from categories A and B were searched. Terms from categories B and C were combined, and five search term combinations from each tobacco brand name were randomly selected using a website providing an algorithm for generating random numbers. To account for differing results from different search engines, three that employed different algorithms were used. The first 200 pro-tobacco-related websites from each search term were recorded. After removal of duplicates, there was a total of 716 websites from which three sites were randomly selected for coding each week for reliability purposes. A coding manual and procedures were designed by a research team, in collaboration with a senior research associate and a clinical psychologist. After training on 12 websites and proving satisfactory reliability, each coder was assigned 15 websites to code per week over the period from November 1999 to May 2000. The content analysis looked for the presence or absence of five features: site category, online purchasing of tobacco products and consumer-awareness information, portrayal of human characters, lifestyle and message

Category A General Smoking Terms	Category B Terms Commonly Associated with Tobacco		Category C Tobacco Companies
Tobacco	Sports	Vacations	Brown & Williamson
Cigarettes	Car racing	Glamour	Philip Morris
Smokeless tobacco	Tennis	Romance	Ligget
Chew tobacco	Rodeo	Woman	RJ Reynolds
Cigars	Celebrities	Fetish	
Pipe tobacco	Movies	Sex	
Snuff	Film	Gambling	
	Freedom	Wine	
	Rights	Cognac	
	Adventure	Beer	
	Travel	Champagne	
	Cruises		

Adapted from Hong & Cody (2002)

Table 5.26 Categories of Search Terms Used to Sample Pro-Tobacco Websites

appeals, and interactive site features. It is therefore recognized that this media is an unmonitored, unregulated source of tobacco marketing targeting young people.

Internal tobacco industry document analysis:

Following the Master Settlement Agreement in the USA and the Health Select Committee's investigation into tobacco companies in the UK, online databases of the tobacco industry's internal documents are available to search[2] (http://www. tobacco-archives.com and http:// www. tobaccopapers.com). The analysis of industry documents has shown that they clearly recognise the power of advertising to retain and recruit smokers, despite their public pronouncements to the contrary (Hastings & McFadyen, 2000; Cummings *et al.*, 2002a). Similarly, documents detailing industry's reactions to "inputs," and their strategies for dealing with them, can be used to measure the effects of tobacco restrictions. For example, a study into how the tobacco industry circumvented Singapore's advertising ban based its findings on internal industry documents (Assunta & Chapman, 2004b). In this study, document collection websites, primarily the Tobacco Archives, were systematically searched using geographic terms and the names of public and private entities relating to Singapore. The resulting documents were then dated, evaluated according to their degree of importance, and a select group were subjected to further analysis. The findings allowed the researchers to examine how the

[2]Marketing search terms for the database should include the following: above the line, advert, below the line, billboard, brand, campaign, coupon, customer, direct mail, email, internet, marketing, mass media, packaging, point-of-purchase, point-of-sale, poster, pricing, product placement, promotion, samples, SMS, target, text message, and website. This list is not exhaustive and care must be taken to search for variations and plurals, possibly by truncation, of the terms above. Brand names should also be included in the strategy (see Cummings *et al.*, 2002a for further search strategies using online databases, and Mekemson & Glantz, 2002 for a sample strategy to locate documents covering tobacco and smoking product placement in movies).

tobacco companies conducted their business in the strict anti-tobacco environment present in the country, and attempted to counter some of the government's tobacco control measures. With knowledge of this kind, it is possible to both tackle the tobacco industry's creative responses to restrictions and monitor the extent to which they are working.

Policy tracking:

Although there has been increased interest in the field of tobacco control policy research, there have been few published accounts of the measurements of the comprehensiveness and strengths of policies (Wakefield & Chaloupka, 1998). A ratings system was developed and implemented which evaluated the extensiveness of state laws restricting youth access to tobacco in the USA (Alciati *et al.*, 1998). State laws were analysed on youth access to tobacco and assigned ratings on nine items: six on tobacco-control provisions, and three on enforcement provisions. For each item, a target was specified reflecting public health objectives. Points were awarded for achieving the target, while criteria for lower ratings were established for situations when the target was not met. Ratings produced by this type of system can, by producing maximum values, indicate that all ideal aspects of a law are in place, facilitate comparison among states (and possibly among countries), permit tracking of changes over time, and make it theoretically possible to relate tobacco control "inputs" to "outputs" (Wakefield & Chaloupka, 1998).

Consumer surveys: the questions to ask

The types of questions that can be used in consumer surveys will be examined, as well as how to identify the issues that should be addressed and developing specific questions to measure them.

Previous studies on the influence of tobacco advertising and marketing can help form a basis for identifying the issues that ought to be examined when measuring the impact of restrictions on marketing. Understanding the relationships between advertising/marketing, and other variables, helps to develop hypotheses about which variables might be expected to be influenced by the elimination of, or severe restrictions on, marketing. The focus here will be on marketing-related proximal and distal variables (see Figure 5.13). Proximal variables are conceptually closest to the restrictions being imposed on marketing communications. First, an assessment must be made of awareness, familiarity, and engagement with specific types of marketing communication to see whether, and to what extent, these lessen when marketing restrictions are imposed. Identifying suitable measures requires familiarity with the content of the marketing restrictions which are to be implemented. This knowledge gives an indication of which marketing practices are going to change or be eliminated, and provides a guide to which measures would be expected to show an impact. For example, in 2003 the UK introduced a comprehensive ban on most forms of tobacco marketing communication, which was implemented in phases from February 2003 until July 2005. It was important to check whether awareness of each prohibited medium, which during the first phase included billboards and press advertising, had reduced. At the same time, it was also useful to measure whether remaining, unrestricted media, which included point-of-sale displays, had increased.

As well as specific media, it is also crucial to monitor the cumulative effect that wide ranging bans can have by disrupting the integrated marketing communications mix. As discussed previously, this is a vital pillar in the industry's attempts to build and maintain evocative brands. It therefore is logical to develop measures of brand salience and image, and monitor how these fair, following policy changes.

Sample questions are drawn from the GYTS and the two ongoing longitudinal studies discussed previously (the ITC Four Country study and the CTCR study). These studies have slightly differing methodologies: the ITC Four Country study is a telephone survey which brings benefits in terms of sampling and ease of respondent access, but limits the

complexity of the questions that can be asked because, for example, it is not possible to use show cards. The CTCR study is a face-to-face, in-home survey which is logistically more difficult, but enables complex questioning procedures; particularly the use of visual aids displaying brand colours and design features. The GYTS is a school-based, self-completion survey of 13-15 year olds, which again limits the complexity of the questions that can be asked.

The studies also target different sub-groups: the ITC consists of a cohort of adult smokers, the CTCR study is conducted with a cross-section of young people aged 11 to 16 years, and the GYTS consists of students aged 13-15 years. Therefore, while some measures may be common, others will be specific to the particular target group. For example, in the ITC study, it makes sense to look at adult smokers' cessation behaviour following marketing restrictions, whereas, with young people in the CTCR study (the majority of whom do not smoke), it is more relevant to look at measures of intention to smoke (see Figure 5.13).

Specific types of marketing communication

Despite a ban on marketing, and thus limited exposure, there can still be significant penetration and continuation of the relationship between marketing and youth smoking (Braverman & Aarø, 2004). Therefore, at the most basic level, there is a need to try and establish how much marketing communication is still getting through. It is difficult to ascertain this, however, without confusing the respondent. Terms like "marketing communications," which are technically correct and capture the generality of the concept, are less likely to be understood than more familiar words like "advertising" or "promotion," which will not capture the breadth of activity that may be involved (see Table 5.23), and may not be consistently interpreted.

Hence, qualitative research played a crucial role in the development of the questionnaire for the CTCR study, ensuring appropriate and comprehensible questioning about a wide range of tobacco marketing activities. While young people could visualise and describe images of conventional advertising (i.e. press, poster, and television adverts), it was much more challenging to get them to think about, and describe, other forms of marketing communications.

The qualitative interviews therefore tried to focus the young respondents' minds on different locations where they might be exposed to tobacco marketing communications, and walk them through various circumstances, asking them to describe any ways that they might see or have their attention drawn to products. For example, they were asked to imagine themselves walking into a shop, and to describe all the things they could see when they approached the door, entered the shop, approached the counter, etc.

In this way the interviews opened their minds to the broader range of marketing practices and encouraged them to describe these in their own terminology. This not only helped with understanding the language and concepts young people use to describe marketing communications, but also revealed the range of promotional activity to which they were aware of being exposed to.

In subsequent focus groups, prompt cards, with descriptions of different forms of tobacco marketing, were developed and presented to respondents, to examine whether or not they could relate to and understand the descriptions. The final stage was to pilot the questions using cognitive interview techniques, whereby respondents were interviewed using the questionnaire and, upon completion, were interviewed to analyse their comprehension of specific questions and their ability to answer them.

The result was the development of questions that described specific tobacco marketing communications in a young person friendly way (see Figure 5.14). Furthermore, because respondents might be interviewed in the presence of a family member, their privacy was protected by presenting the various descriptions on prompt cards, so they could express their answers confidentially.

Figure 5.15 shows how the ITC attempted to gain an overall measure of awareness of tobacco marketing using very general lay terminology. Whereas the CTCR

question did not impose a time frame, the ITC study tried to limit recall to the previous six months to help participants focus their attention on a specific and manageable period. This is less than ideal; it could, for instance, pick up non-marketing influences such as peer smoking or be interpreted differently by respondents. Nonetheless, it does help to start putting together a picture of what may be happening. When asked alongside other more specific questions, it provides a useful gauge for the amount of pro-smoking messages that are being perceived. Furthermore, it is likely to provide a general measure of tobacco marketing, as it has been argued that advertising can affect behaviour even if an advert is not actively processed and respondents cannot recall seeing it (Shapiro *et al.*, 1997).

Figure 5.16 looks at specific media and measures how successfully any controls are working by examining awareness of communications in each of these. The media included in this question can be varied to suit the jurisdiction (e.g. in the UK, where television advertising for tobacco products has been forbidden for nearly 20 years, this option may be omitted).

The GYTS survey takes a slightly different approach. It focuses on specific media and asks young people to rate the amount they have seen within a short prior time period of one month (Figure 5.17). While these questions are likely to sufficiently discriminate between those who do and do not recall each form of advertising, there is the possibility of some ambiguity over the amount recalled. Response categories of "a lot" and "a few" may be too ambiguous to appropriately distinguish between different amounts recalled; one respondent's perception of "a lot" may be another's perception of "a few".

Other forms of tobacco marketing communications need to be addressed with separate questions; sports and event sponsorship need careful consideration. In the ITC survey it was important to try and distinguish between overt brand sponsorship (e.g. Marlboro or Formula 1) from more covert corporate social responsibility (e.g. the British American Tobacco Company's support for good causes, such as farming methods in Malawi). Therefore, a rather complicated set of questions were asked here (Figure 5.18).

In the ITC survey, respondents found it difficult to answer this bank of questions, so it may be preferable to use the slightly simpler version presented in Figure 5.19. This is a classic example of the dilemma faced by questionnaire designers: how to reflect the complexities of the real world by phrasing accurate questions that do not cause confusion (for a detailed discussion about issues related to question wording see Oppenheim, 1992).

The CTCR study was also interested in which sports or events young people associated with tobacco and, where possible, the brands they connected with these (Figure 5.20).

The GYTS measures awareness of cigarette brands on TV, including those within coverage of sporting events (see Figure 5.21). This is likely to provide a measure of overall awareness of cigarette brands on television, but does not specifically measure awareness of sports sponsorship. Again, the response categories rate frequency, which may give rise to ambiguity.

The remaining form of marketing communication is, rather confusingly, referred to as a "promotion." This can come in many guises: from money-off coupons to free samples, as illustrated in Table 5.24. All these variants need to be covered. An extra complexity is the need to measure not just awareness of these activities, but participation in them (e.g. have people taken advantage of price promotions, as well as hearing about them). The ITC study drew on knowledge gained from the CTCR study and also subdivided promotions down into specific descriptions of marketing (Figure 5.22).

Measuring branding

Branding is a traditional advertising method used to create a response from a target audience based on cumulative impressions and positive reinforcement. At one level, measuring branding is no more complex than measuring individual marketing communications, and simple measures can be constructed to determine spontaneous and prompted awareness of different brands

I'm going to show you some cards (SHOWCARDS 10-26) with descriptions of some other ways that companies might try to attract attention to cigarettes. For each one can you tell me if you have seen anything like this.

(Answer categories were: Yes; No; or Don't Know. For each marketing type responders were aware of, they were asked to say which make or brand it was connected with.)

a. SHOWCARD 10 Adverts for cigarettes on large posters or billboards in the street
b. SHOWCARD 11 Adverts for cigarettes in newspapers or magazines
c. SHOWCARD 12 Signs or posters about cigarettes in shops or on shopfronts:
   on shop windows
   on shop doors
   on cigarette display units inside shops
   on clocks inside shops
   on staff aprons or overalls
   on signing mats inside shops
   some other sign or poster about cigarettes (in shops or on shopfronts)
d. SHOWCARD 13 Free trial cigarettes being given out or offers to send away for free cigarettes
e. SHOWCARD 14 Free gifts from the shop keeper when people buy cigarettes
f. SHOWCARD 15 Free gifts when people save coupons or tokens from inside cigarette packs
g. SHOWCARD 16 Free gifts when people save parts of cigarette packs (eg. pack fronts)
h. SHOWCARD 17 Free gifts, showing cigarette brand logos, being given out at events such as concerts, festivals or sports events
i. SHOWCARD 18 Special price offers for cigarettes
j. SHOWCARD 19 Promotional mail, from cigarette companies, being delivered to people's homes
k. SHOWCARD 20 Clothing or other items with cigarette brand names or logos on them
l. SHOWCARD 21 Competitions or prize draws linked to cigarettes
m. SHOWCARD 22 Famous people, in films or on TV, with a particular make or brand of cigarettes
n. SHOWCARD 23 New pack design or size
o. SHOWCARD 24 Internet sites promoting cigarettes or smoking (do **not** include anti-smoking sites)
p. SHOWCARD 25 Email messages or mobile phone text messages promoting cigarettes or smoking (do **not** include anti-smoking messages)
q. SHOWCARD 26 Leaflets, notes or information inserted in cigarette packs
r. NO SHOWCARD Have you come across any other ways that companies try to attract attention to cigarettes?

Figure 5.14 Question assessing awareness and involvement in tobacco promotions

Centre for Tobacco Control Research (CTCR) Ad-ban study (University of Strathclyde)

Thinking about everything that happens around you, in the last 6 months how often have you noticed things that promote smoking?

01 – Never
02 – Rarely
03 – Sometimes
04 – Often
05 – Very Often

Figure 5.15 General Measurement of Pro-Smoking Messages in the International Tobacco Control Policy Evaluation Study

Now I want to ask you about tobacco advertising. In the last 6 months, have you noticed cigarettes or tobacco products being advertised in any of the following places?

(Read out each statement)
01 – Yes
02 – No

a. On television
b. On radio
c. At the [cinema/movie theatre], before or after the [film/movie]
d. On posters or billboards
e. In newspapers or magazines
f. On [shop store] windows or inside [shops/stores] where you buy tobacco
g. Other

Figure 5.16 Measuring Awareness of Tobacco Ads in Specific Media in the International Tobacco Control Policy Evaluation Study

(Figure 5.23). The latter of course requires visual prompts depicting a selection of brands. However, there is the need to delve deeper and assess not just the ability to recall brands, with or without prompting, but familiarity and engagement with them. The former can be done by checking if respondents can complete partially masked brand examples, as in Figure 5.24.

Deeper engagement ventures into the rather illusive area of "brand image:" the emotional associations and feelings that are attached to marques, such as for Marlboro or Benson & Hedges. As noted earlier, the tobacco industry goes to great lengths and expense to create evocative images for their brands, and arguably a key task of tobacco control in general, and marketing restrictions in particular, is to undermine them. Measuring the results is tricky; this type of complexity lends itself more readily to qualitative methods than quantitative ones. Nonetheless, questionnaires can be used successfully to tackle the problem. Figure 5.25 illustrates how semantic scales can help unravel dimensions like popularity, appeal to specific sub-groups, and masculinity. Rating, ranking, and "pick-any" (in which respondents are asked which brand(s), if any, they associate with a series of attributes) measures of brand image associations have been reported to be comparable (Driesener & Romaniuk, 2006).

Conclusions

This section has explained what is meant by tobacco marketing communications, that they do influence tobacco consumption, especially by the young, and that it is therefore crucial to instigate controls and measure their effectiveness. It has been shown that this can best be done by monitoring a range of distal and proximal variables using consumer surveys.

Consumer surveys are further enhanced when surveillance systems are put in place to monitor changes in tobacco marketing activity following restrictions. This helps in contextualising the findings and

During the past 30 days (one month), how many advertisements for cigarettes have you seen on billboards?

a. A lot
b. A few
c. None

During the past 30 days (one month), how many advertisements or promotions for cigarettes have you seen in newspapers or magazines?

a. A lot
b. A few
c. None

Figure 5.17 Measuring Awareness of Tobacco Ads in Specific Media in the Global Youth Tobacco Survey

In the last 6 months, have you seen any advertising by tobacco companies that is NOT promoting particular products or brands, but the COMPANY itself?

01 – Yes
02 – No

Still thinking about the last 6 months, have you seen or heard about any sport or sporting event that is sponsored by or connected with BRANDS of cigarettes?

01 – Yes
02 – No

In the last 6 months, have you seen or heard about any sport or sporting event that is sponsored by or connected with tobacco COMPANIES?

01 – Yes
02 – No

In the last 6 months, have you seen or heard about any music, theatre, art, or fashion events that are sponsored by or connected with BRANDS of cigarettes?

01 – Yes
02 – No

In the last 6 months, have you seen or heard about any music, theatre, art, or fashion events that are sponsored by or connected with tobacco COMPANIES?

01 – Yes
02 – No

Figure 5.18 Measuring Tobacco Sponsorship the Hard Way in the International Tobacco Control Policy Evaluation Study

In the last 6 months, have you seen or heard about any sport or sporting event that is sponsored by or connected with a tobacco company or brand?

01 – Yes
02 – No

In the last 6 months, have you seen or heard about any music, theatre, art, or fashion events that are sponsored by or connected with a tobacco company or brand?

01 – Yes
02 – No

Figure 5.19 A Simpler Way of Measuring Tobacco Sponsorship in the International Tobacco Control Policy Evaluation Study

interpreting any changes or lack of expected changes.

The consumer surveys need to take baseline measures prior to any changes in marketing restrictions, and several follow-up surveys over a period of years to monitor short-term and longer-term effects. The length of follow-up will partly be dictated by the implementation time table of the restrictions. For example, in the UK their ban was implemented in phases, making it conducive to conducting follow-up surveys after each phase.

The final subsection examined specific questions that have been successfully used to do this monitoring and showed how particular questions will vary depending on the target group and the administration mode.

In applying the methodologies discussed here, however, it is important to recognise that the precise wording of questions will vary according to the sample being interviewed. Before going into the field, therefore, it is crucial to conduct a thorough pilot study. This should include qualitative work to check matters of content and language, and quantitative research to check understanding and feasibility.

Can you think of any sports or games that are sponsored by or connected with any makes or brands of cigarettes?

FOR EACH SPORT or GAME MENTIONED, ASK: What make(s) or brand(s) is it connected with?

PROBE FOR SPORT/GAME AND MAKE(S)/BRAND(S)
REPEAT FOR MAXIMUM OF 6 SPORTS/GAMES

Sport or Game	Make(s) or Brand(s)
1. ……	…….....
2.……	…….....
3. ……	…….
4. ……	……..... .
5. ……	…….....
6.……	. …….....

Can you think of any other events or shows that are sponsored by or connected with any makes or brands of cigarettes?

FOR EACH EVENT or SHOW MENTIONED, ASK: What make(s) or brand(s) is it connected with?

PROBE FOR EVENT? SHOW AND MAKE(S)/BRAND(S)
REPEAT FOR MAXIMUM OF 6 SPORTS/GAMES

Event or Show	Make(s) or Brand(s)
1. ……	…….....
2. ……...	…….....
3. …….....	…….....
4. ……...	…….....
5. ……...	…….....
6.……	…….....

Figure 5.20 Measuring Awareness of Tobacco Sponsorship Among Young People in the Centre for Tobacco Control Resarch (CTCR) study at the University of Strathclyde

During the past 30 days (one month), when you have watched sports events or other programmes on TV how often did you see cigarette brand names?

a. I never watch TV
b. A lot
c. Sometimes
d. Never

Figure 5.21 Measuring Sponsorship in the Global Youth Tobacco Survey

In the last 6 months, have you **noticed** any of the following types of tobacco promotion:
READ OUT EACH STATEMENT

01 – YES
02 – NO

a. Free samples of cigarettes. **If yes**: Have **you** received free samples of cigarettes?
b. Special price offers for cigarettes. If **yes**: have **you** used special price offers?
c. Free gifts or special discount offers on other products when buying cigarettes?
d. (IF YES) Were these free gifts or special discounts from:

1. the shop-keeper when buying cigarettes
2. you or someone else saving coupons or tokens from inside cigarette packs
3. you or someone else saving parts of cigarette packs (e.g. pack fronts)
4. free gifts showing cigarette brand logos, given out at events such as concerts, festivals or sports events

If yes to any of the above ask: Have you personally received such gifts?

e. Email messages promoting cigarettes or tobacco products. **If yes**: Have you received promotional email messages?
f. Mobile phone text messages promoting cigarettes or tobacco products. **If yes:** Have you received mobile phone text messages…
g. Mail promoting cigarettes or tobacco products. **If yes**: Have you received…..
h. Clothing or other items with a cigarette brand name or logo. If yes: have you received….
i. Competitions linked to cigarettes. **If yes:** have you participated in any competitions linked to cigarettes?
j. Internet sites promoting cigarettes or tobacco products. **If yes**: Have you visited any internet sites…..
k. Leaflets promoting cigarettes or tobacco products. **If yes**: Have you received any leaflets …..
l. Signs or posters or branded items in bars, pubs or clubs

Figure 5.22 Measuring Awareness and Involvement in Tobacco Promotions in the International Tobacco Control Policy Evaluation Study

Can you tell me the names of as many makes or brands of cigarettes that you have either seen or heard of:

Record up to a maximum of 10 ……..……

Question: Now can you tell me whether you have seen any of these makes before?

VISUAL PROMPTS 6-10
POINT TO EACH PROMPT ONE AT A TIME
FOR EACH ONE ASK: Have you ever seen this one?

Visual prompt	Yes	No	Don't Know
6. Windsor Blue			
7. Berkeley			
8. Benson & Hedges			
9. Lambert & Butler			
10. Marlboro			

Figure 5.23 Measurement of Brand Awareness in the Centre for Tobacco Control Research (CTCR) study at the Univ ersity of Strathclyde

I'm going to show you some packets of cigarettes that have the name covered up on them. For each one I'd like you to tell me what make or brand you think it is. Please don't worry if you don't know the make or brand.

SHOW VISUAL PROMPTS

This brand is very popular with people my age						This brand is very unpopular with people my age DK
Benson & Hedges	1	2	3	4	5	6
Lambert & Butler	1	2	3	4	5	6
Marlboro	1	2	3	4	5	6

You never see this brand in shops around here						You always see this brand in shops around here DK
Benson & Hedges	1	2	3	4	5	6
Lambert & Butler	1	2	3	4	5	6
Marlboro	1	2	3	4	5	6

Figure 5.24 Measurement of Brand Familiarity in the Centre for Tobacco Control Research (CTCR) at the University of Strathclyde

Most smokers smoke this brand						Few smokers smoke this brand
						DK
Benson & Hedges	1	2	3	4	5	6
Lambert & Butler	1	2	3	4	5	6
Marlboro	1	2	3	4	5	6
Attractive looking brand						Unattractive looking brand
						DK
Benson & Hedges	1	2	3	4	5	6
Lambert & Butler	1	2	3	4	5	6
Marlboro	1	2	3	4	5	6
Female brand						Male brand
						DK
Benson & Hedges	1	2	3	4	5	6
Lambert & Butler	1	2	3	4	5	6
Marlboro	1	2	3	4	5	6

Figure 5.24 Measurement of Brand Familiarity in the Centre for Tobacco Control Research (CTCR) at the University of Strathclyde

5.5 Measures to evaluate the effectiveness of tobacco product labelling policies

Background

The cigarette package serves as the cornerstone of tobacco marketing and advertising campaigns (Slade, 1997; Pollay, 2001). Package design helps to reinforce brand imagery communicated through other media and plays a central role in retail marketing. The importance of cigarette packaging only increases as other forms of marketing are restricted, as indicated in the following quote from a Philip Morris executive: "Our final communication vehicle with our smoker is the pack itself. In the absence of any other marketing messages, our packaging...is the sole communicator of our brand essence. Put another way—when you don't have anything else—our packaging is our marketing." (Alechnowicz & Chapman, 2004).

Governments in many jurisdictions have begun to apply greater restrictions on tobacco labelling. As much as half of the package is now used by regulators to communicate the health effects of smoking. Governments have also begun to prohibit packaging elements that are deemed to be misleading to smokers. As a consequence, labelling policies have begun to alter the traditional appearance of the cigarette package.

The importance of tobacco labelling policies is highlighted in Article 11 of the WHO FCTC (WHO, 2003). Article 11 sets international standards for packaging and labelling of tobacco products in three broad categories: 1) mandatory health warnings; 2) restrictions on brand descriptors, such as the use of "light" and "mild"; and 3) information on cigarette contents and emissions[1] (Figure 5.26).

Health warning labelling

Cigarette packages in the vast majority of countries carry a health warning (Aftab *et al.,* 1999). However, the position, size, and general strength of these warnings vary considerably across jurisdictions. FCTC Article 11 requires that package health warnings must cover at least 30% of the package surface and be "large, clear, visible, and legible" (WHO, 2003). Beyond these minimum requirements, Article 11 also states that warnings "should" cover 50% or more of a package's principle surfaces, and "may" be in the form of pictures.

To date, at least eight countries have implemented picture-based health warnings that meet the FCTC's "recommended" standard (see Figure 5.27). A number of other jurisdictions, including the European Union, have recently implemented prominent text warnings which meet the minimum FCTC standard. More obscure text warnings remain in many other markets, including the USA, China, and Russia.

Constituents & emissions labelling

There is general agreement that tobacco packaging should include at least minimal information about some of the hazardous and addictive constituents in tobacco and tobacco smoke. FCTC Article 11 requires that packages contain "information on relevant constituents and emissions of tobacco

[1]Tobacco labelling policies apply to a broad range of tobacco products, including a range of combustible products, such as cigars, and the packaging of loose or "fine cut" tobacco, as well as non-combustible tobacco products. However, much of this section will focus on labelling policies for factory-made, "pre-packaged" cigarettes given that they are the primary target of labelling policies, and the area in which most research has been conducted. Labelling policies for other types of products will be described briefly in a separate section to follow.

1. Each Party shall, within a period of three years after entry into force of this Convention for that Party, adopt and implement, in accordance with its national law, effective measures to ensure that:

(a) Tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions, including any term, descriptor, trademark, figurative or any other sign that directly or indirectly creates the false impression that a particular tobacco product is less harmful than other tobacco products. These may include terms such as “low tar”, “light”, “ultra-light”, or “mild”; and

(b) Each unit packet and package of tobacco products and any outside packaging and labelling of such products also carry health warnings describing the harmful effects of tobacco use, and may include other appropriate messages. These warnings and messages:

(i) shall be approved by the competent national authority,
(ii) shall be rotating,
(iii) shall be large, clear, visible and legible,
(iv) should be 50% or more of the principal display areas but shall be no less than 30% of the principal display areas,
(v) may be in the form of or include pictures or pictograms.

2. Each unit packet and package of tobacco products and any outside packaging and labelling of such products shall, in addition to the warnings specified in paragraph 1(b) of this Article, contain information on relevant constituents and emissions of tobacco products as defined by national authorities.

3. Each Party shall require that the warnings and other textual information specified in paragraphs 1(b) and paragraph 2 of this Article will appear on each unit packet and package of tobacco products and any outside packaging and labelling of such products in its principal language or languages.

4. For the purposes of this Article, the term “outside packaging and labelling” in relation to tobacco products applies to any packaging and labelling used in the retail sale of the product.

WHO (2003)

Figure 5.26 WHO FCTC Article 11: *Packaging and labelling of tobacco products*

products as defined by national authorities.” At present, however, national authorities have taken much different approaches to labelling constituents and emissions, and there remains considerable disagreement regarding what should be considered “relevant” information.

The current regulatory practice in many jurisdictions is to require manufacturers to print levels for three emissions in the mainstream smoke: tar, nicotine, and carbon monoxide (CO). Emission levels are generated by machine-smoking cigarettes according to a standard set of puffing conditions; typically the International Standards Organization (ISO) method, which serves as the current international standard. However, in light of research indicating that the tar and nicotine levels generated under the ISO testing method are unrelated to individual levels of exposure or risk (Burns *et al.*, 2001), there are growing calls from within the tobacco control community for the ISO numbers to be removed from packages. Some jurisdictions have supplemented the ISO numbers with additional emission information. For example, Canada increased the list of emissions that must be reported

Figure 5.27 Picture-based warnings

(by adding benzene, formaldehyde, and hydrogen cyanide to tar, nicotine, and CO), and requires a second set of numbers from a more intensive machine smoking method for each emission (Figure 5.28). Other jurisdictions have replaced quantitative emission values with descriptive, non-numerical information on hazardous emissions and toxicants. A consensus has yet to emerge on "best practices" for this area of tobacco labelling policy.

Brand descriptor labelling

Tobacco manufacturers incorporate a variety of common terms into the names of their cigarette brands. Words such as "light" and "mild" are ostensibly used to denote flavour and taste; however, "light" and "mild" brands are often promoted as "healthier" products and are typically applied to brands that generate lower machine levels of tar (Pollay, 2001; Pollay & Dewhirst, 2002). Not surprisingly, "light" and "mild" brands are perceived by many consumers to deliver less tar and lower risk than "regular" or "full flavour" varieties despite evidence to the contrary (Ashley *et al.,* 2001; Shiffman *et al.*, 2001).

A growing number of jurisdictions, including Brazil and the European Union, have prohibited the use of "light" and "mild" on packages. Similar prohibitions are proposed in FCTC Article 11: "tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions, including any term, descriptor, trademark, figurative or any other sign that directly or indirectly creates the false impression that a particular tobacco product is less harmful than other tobacco products. These may include terms such as "low tar," "light," "ultra-light," or "mild". Although there is evidence to suggest that other packaging elements, such as the use of colour, may also create misleading perceptions of risk (Wakefield *et al.*, 2002), "light" and "mild" descriptors are the only packaging elements to be restricted to date.

Methodological issues in evaluating tobacco labelling policies

Evaluating tobacco labelling policies presents several unique challenges; this section reviews some of the principal methodological and analytical considerations.

"Alternative" tobacco products:

Labelling policies have generally been designed with factory-made, pre-packaged cigarettes in mind. However, a substantial proportion of tobacco users throughout the world use tobacco products that are either packaged in a different way, or have no manufactured packaging at all. This has important implications for patterns of exposure to health warnings. For example, consumers who buy loose or fine-cut tobacco, without any manufactured packaging, may not be exposed to product health warnings. Even consumers who buy fine-cut tobacco, sold in government-mandated packaging, will have different patterns of exposure than those who smoke manufactured cigarettes, and who are likely to be exposed to the warnings each time they reach for the package. As a result, studies conducted in markets with a considerable proportion of fine-cut tobacco sales, such as the United Kingdom, New Zealand, and Thailand, may need to stratify for fine-cut versus manufactured or mixed use. Smuggled or contraband cigarettes may also alter patterns of exposure in cases when the contraband product is not manufactured to the same labelling specifications.

Issues in attribution: dealing with multiple sources of health information:

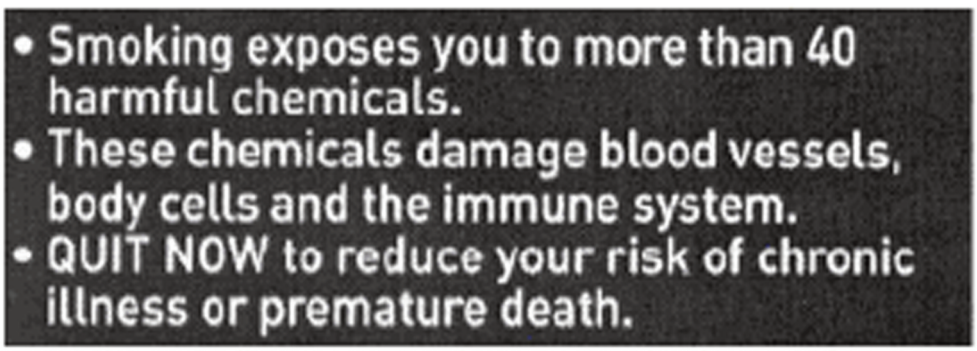
Health behaviours with multiple determinants present a challenge to policy evaluation. The problem of attribution is particularly acute for health warning labels. First, labelling policies are often implemented simultaneously with other tobacco control measures, including increases in taxation and smoke-free policies. As a result, it is difficult to isolate the effect of an individual policy on overall prevalence. Second, many of the specific themes and messages in labelling policies are communicated through other sources.

European Union (United Kingdom): Three ISO emissions

"Toxic emissions / unit:" "Tar" 14 - 35 mg,
Nicotine 1.1 - 2.7 mg, Carbon monoxide 14 - 30 mg,
Formaldehyde 0.055 - 0.14 mg,
Hydrogen cyanide 0.14 - 0.36 mg, Benzene 0.043 - 0.097 mg

"Émissions toxiques / unité:" "Goudron" 14 - 35 mg,
Nicotine 1,1 - 2,7 mg, Monoxyde de carbone 14 - 30 mg,
Formaldéhyde 0,055 - 0,14 mg,
Acide cyanhydrique 0,14 - 0,36 mg, Benzène 0,043 - 0,097 mg

Canada: Six ISO emissions and 'Health Candada Intense' emission



Australia: Descriptive information

Figure 5.28 Constituent labelling policies in the European Union, Canada and Australia

Mass media campaigns and health professionals often target the same health effects, particularly with regards to common diseases such as cancer and cardiovascular disease. The impact of package-based labelling policies may also be confounded with health warnings in other settings. Various jurisdictions require health warnings in retail outlets and warnings on print advertisements for tobacco products. Third, perceptions of risk and health knowledge are influenced by an inter-related set of factors at the individual, social, and environmental level. Few studies are able to measure more than a small number of these factors within a single study and none can fully isolate the contributions of each. These realities underscore the importance of the methodological features described in Section 2.1. In addition, environmental scans of other mass media campaigns and policy interventions can provide important context.

"Wear-out" and impact over time:

It is widely accepted that the salience of advertising and health communications is typically greatest upon first exposure (Bornstein, 1989; Henderson, 2000). The initial impact of comprehensive labelling policies, such as the introduction of large graphic warnings on packages, is often magnified by media coverage. As a result, measures of effectiveness are likely to be strongly associated with the implementation date. This has implications for regulators in terms of ensuring periodic changes to the warnings, as well as studies that compare labelling policies across jurisdictions. For example, a recent study found that new text-based warnings, introduced in the United Kingdom in 2003, were considerably more likely to be noticed than Australian text-based warnings, which were only slightly smaller, but had been in place for more than eight years at the time of the survey (Bornstein, 1989). Ideally, labelling policies should be evaluated at similar post-implementation dates; at the least, differences in follow-up periods should be clearly noted and taken into account when interpreting findings.

There is preliminary evidence to suggest that not all measures of effectiveness decline at the same rate over time. "Proximal" measures of salience, such as noticing warnings, may erode more quickly than "distal" measures, such as reporting that health warnings motivate quitting and increase thoughts about the health risks of smoking (Hammond *et al.*, 2007a). It is even plausible that for some smokers the impact of health warnings could increase over time. For example, the cessation and telephone quitline information included in many health warnings may only become relevant to smokers as they contemplate quitting. In a population-based survey, however, the ebb and flow among individuals will balance out, and one would still anticipate decreases in measures of effectiveness over time.

Youth:

One policy-relevant question concerns the impact of warning labels in reducing youth uptake. Evaluating the impact of health warnings among youth during periods of smoking initiation requires a different conceptual approach. Given that the cigarette package serves as the medium for labelling policies, consumption levels may be positively associated with knowledge of the warning labels. In other words, individuals who smoke 20 cigarettes a day will be exposed to the warnings more frequently than individuals who smoke less than daily. Furthermore, "occasional" youth smokers are less likely to buy their own package, reducing the likelihood of exposure to warning labels, compared to more regular smokers who are more likely to buy their own package (Leatherdale, 2005). As a result, individuals who smoke more frequently are more likely to recall the content, location, and other aspects of labelling policies, a counter-intuitive association at first glance (Robinson & Killen, 1997).

A second issue concerns longitudinal studies that use measures of exposure or knowledge as predictors of future smoking behaviour among youth. During youth and young adulthood, the rate of smoking undergoes significant increases. As youth smoking behaviour increases, so too will their exposure to the package and their knowledge of the warnings. Thus, whereas a negative association

between exposure and future smoking behaviour may be expected for anti-smoking campaigns in other media, this is not the case for warning labels.

Failure to account for the somewhat counter-intuitive association between smoking and exposure to health warnings can result in misleading interpretations of data. For example, one study characterized an association between increased smoking and increased knowledge of health warnings as "paradoxical," and also found evidence that US health warnings were ineffective (Robinson & Killen, 1997). This may have been the case; however, without a comparison group, the authors had no way of knowing whether the increases in smoking behaviour were greater, less, or no different than they would have been if no warnings or more comprehensive warnings had been implemented. It may be, for example, that fewer youth initiated smoking than would have otherwise occurred without the health warnings. In fact, this was the pattern reported in a longitudinal study comparing changes in youth smoking in Canada and the USA following the introduction of graphic warning labels on Canadian packages. Smoking rates and knowledge of the warnings rose among Canadian youth as they aged; however, the increase in smoking was significantly less than among US adolescents and the increase in knowledge of the warnings considerably greater (Fong *et al.*, 2002). Overall, this study underscores the importance of suitable research designs and appropriate interpretations of the data when evaluating warning labels among youth.

Evaluation of individual messages & content:

Beyond the question of whether health warnings are generally effective, there is a growing body of research on the individual elements of a warning. These elements can broadly be categorized in terms of design and content components. To date, much of the research has focused on important design elements including the size, position, and use of pictures on the package (Strahan *et al.*, 2002). In contrast, relatively few studies have examined the content of individual messages.

Population-based surveys that compare labelling policies across time or jurisdictions are somewhat ill-suited to the task of evaluating individual warnings. Policies typically differ on more than one dimension, and policy changes typically involve increases in the size, number, position, and type of information presented in each warning. Evaluating individual components or messages becomes more complicated as the number of warnings and complexity of information increases; it is far easier to evaluate the effectiveness of a single warning through survey-based research than to evaluate the content of 16 individual warnings.

When assessing the impact of individual warnings, it is also important to consider that many health warnings are tailored to particular sub-groups of smokers. Warnings on the risks of smoking while pregnant, for example, have little relevance for older males. Thus, it is conceivable that some warnings may perform very well among sub-groups who comprise the target audience, but relatively poorly among the population as a whole. As a consequence, survey measures may need to be adapted and the findings may need to be stratified among relevant sub-groups. One might expect the tailoring of warnings to increase, as the use of picture-based warnings increase, along with the typical number of rotating warnings in a given jurisdiction.

In general, population-based surveys may be most appropriate for identifying the overall effectiveness of a set of health warnings. However, the task of evaluating the content of individual warnings is best suited to experimental or qualitative designs, in which the content and design can be systematically varied.

Geographic & cultural differences:

Very little research has examined potential geographic and cultural differences in the effectiveness of health warnings. Although the fundamental principles underlying the effectiveness of warnings are unlikely to vary across cultures and regions, the effectiveness of individual messages may indeed perform differently. First, smokers in different parts of the world have different levels of existing health knowledge. This has implications

for the type of messages to be included in warnings. For example, Australian smokers may have a relatively higher level of health literacy than smokers in other regions, which may account for the decision to include a warning for "peripheral vascular disease" on packages. Picture-based warnings may be particularly important in populations with lower literacy rates (CRÉATEC, 2003). In addition, the images used in one jurisdiction may not be equally effective in another. For example, several of the picture-based warnings that appear on Venezuelan and Uruguayan packages, and elsewhere, use symbols that may be culturally specific. Finally, similar sets of warnings may be more effective in areas where smokers have relatively little access to anti-smoking information from mass media or health professionals. Few of these issues have been addressed to date; however, they are likely to gain prominence as a growing number of jurisdictions in Asia, Africa, and the Middle East enhance their labelling policies to meet Article 11, and must rely on an evidence base that derives from relatively few Western and Latin American countries.

Evaluating the removal of information:

Unlike other labelling policies, restrictions on brand descriptors result in the removal, rather than the provision of, information. This presents a challenge to evaluation, particularly when the information being removed is used as a brand identifier. In the case of bans on the use of "light" and "mild," the terminology that was previously used to identify a class of products no longer exists. Smokers may retain the same misleading perceptions of these products after the terms have been prohibited, but survey measures can no longer refer to "light" or "mild" cigarettes in the same way as in the past. Therefore, survey measures must be designed so that the wording and meaning of questions remains constant before and after the removal of these terms. This creative challenge is only now being confronted by researchers with the recent advent of "light" and "mild" prohibitions. One approach, discussed later in this section, is to make the respondents' "own brand" the referent for questions.

Another implication of the "removal" of brand information is that the beliefs associated with "light" and "mild" cigarettes are likely to persist for some time after the descriptors disappear from packages. This situation is similar to advertising, promotion, and sponsorship bans; one should not expect beliefs to change immediately upon the implementation of the policy, but more gradually over time. Indeed, anecdotal evidence suggests that many retailers and consumers continue to use the terms "light" and "mild" well after their removal. Other packaging elements and aspect of cigarette design may also reinforce the same beliefs and perceptions as the "light" and "mild" descriptors. These considerations are important in terms of how the data are interpreted and how the "effectiveness" of light and mild policies is conceptualized.

Defining misleading descriptors:

There is widespread confusion among both consumers and many within the tobacco control community regarding several key terms relevant to labelling policy. Many fail to make the distinction between "light" and "mild" and "low tar." Whereas "light" and "mild" are terms used in the name of a brand, strictly speaking "low tar" refers to the emission levels under machine testing. Although there is a very strong correlation between the two (manufactures often attach "light" and "mild" descriptors to brands that generate lower tar levels under the ISO smoking machine), one can have a "light" cigarette that does not generate "low tar" levels and vice versa. Strictly speaking, in jurisdictions with bans, "light" and "mild" cigarettes do not exist, whereas "low tar" cigarettes do. To complicate matters further, the terms "light" and "mild" can also be used to refer to sensory properties of a cigarette. Thus, smokers may still retain the concept of a cigarette as "light" or "mild" even in the absence of a brand descriptor. Given the potential for confusion, survey measures should be explicit about the intended meaning of these terms and should avoid using them interchangeably. This becomes

apparent when measuring these concepts in jurisdictions where the "light" and "mild" brand descriptors have been removed.

Measures

This section provides an overview of the key constructs and individual measures that have been used to assess labelling policies. The constructs range from the extent to which labelling policies are noticed and processed, the extent to which they alter key beliefs (such as levels of health knowledge), to their impact upon downstream behavioural outcomes. These measures can be organised within a conceptual model, as illustrated in Figure 5.29. Other psychosocial variables, such as social norms and beliefs about the tobacco industry, could also be added to this model, but have been excluded in the interest of brevity. The following sub-section begins with a review of quantitative measures, followed by qualitative measures, and a brief discussion of the role of industry documents (Tables 5.27-5.39; see also Appendices 9 and 10).

Measures of labelling salience and processing:

Health warnings must be cognitively processed to be effective. The extent to which information is processed or elaborated upon has been shown to be the most important determinant of memory and attitude change in response to new information (Anderson, 1990). A number of measures have been developed to assess cognitive processing of health warnings as a general indicator of their salience. These measures range from more "shallow" measures of processing, such as a general awareness of warnings, to "deeper" measures of processing, including reading the warnings and thinking about them when they are not in sight (Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond

Construct	Noticing Health Warnings
Measure	"In the last month, how often, if at all, have you noticed the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often)
Sources	Hammond *et al*., 2006a; Hammond *et al*., 2007a
Validity	The time reference varies across different versions: some questions include no time reference ("How often do you notice…"), whereas others refer to the "last month" or "last three months." The response categories also vary and are often collapsed into a smaller number of categories in analysis. The basic question can also be asked within the context of noticing other forms of anti-tobacco media (e.g. "In the last 6 months, have you noticed advertising or information that talks about the dangers of smoking, or encourages quitting in any of the following places? (Yes, No to a list of 9 media channels, including on *cigarette packages*)).
Comments	Overall, a straightforward measure of the salience and processing of warnings that should be considered within the core set of variables to assess health warnings. As close to a "gold standard" in this domain as exists. Using the same wording to ask about salience of other media channels provides a useful comparative index for the salience of various health information channels. A recommended and essential measure for evaluating health warnings.

Table 5.27 Essential Measure of Labelling Salience and Processing

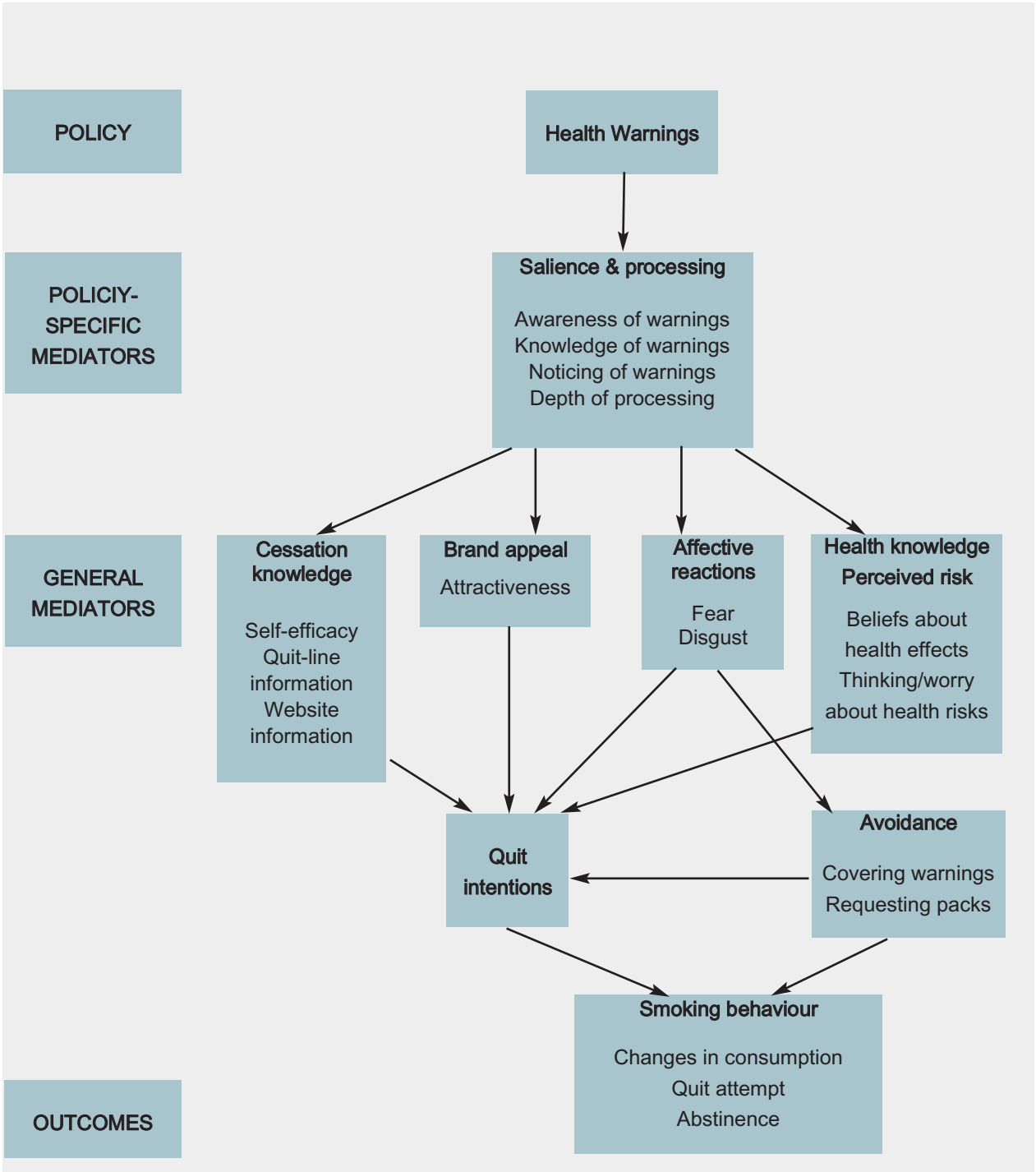


Figure 5.29 Conceptual framework for the evaluation of health warning policies

Construct	(a) General Awareness
Measure	"Have you seen health warnings on cigarette packages?" (Yes, No)
Sources	Borland & Hill, 1997a; Health Canada, 2005
Validity	Good face validity; associated with policy strength.
Variations	Response categories are consistent across measures. Alternative wordings include: "Are you aware of health warnings on cigarette packages?" and "Are there health warnings on packages?" Some questions refer specifically to the release of new warnings. For instance, "Are you aware of any recent changes to health warnings on cigarette packs?" and "Have you noticed any changes to the health warnings on cigarette packages since [date]?"
Comments	Provides an overall measure of general awareness. Limited value in examining changes and comparing across jurisdictions, given almost universal awareness among smokers. Most useful for examining policy implementation and rollout when the question makes reference to "new" warnings, or in jurisdictions with very weak health warnings and no previous research. Overall, an informative measure, but only recommended under these circumstances.
Construct	**(b) Reading/Looking Closely at Health Warnings**
Measure	"In the last month, how often, if at all, have you read or looked closely at the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often)
Sources	Hammond *et al*., 2006a; Hammond *et al*., 2007a
Validity	Face validity; good convergent validity with other measures; good predictive validity for strength of policy and motivation to quit smoking.
Variations	The time reference varies across different versions. Also, some versions refer to *reading*, other versions use broader language, such as *looking closely,* and some versions include both terms. *Looking closely* may be more appropriate for pictorial warnings.
Comments	Strong correlation with noticing, but conceptualized as a "deeper" measure of processing. Overall, a recommended and important, but not essential, measure of salience and processing that may be particularly relevant for textual aspects of warnings.
Construct	**(c) Discussing the Health Warnings With Others**
Measure	"In the last month, how often, if at all, have you talked about the health warning with others?" (Never, Rarely, Sometimes, Often, Very Often)
Source	Hammond *et al*., 2003

Table 5.28 Additional Measures of Labelling Salience and Processing

Validity	Good face validity, convergent validity, and predictive validity for motivation to quit and future smoking behaviour when included as part of a composite measure.
Variations	Variations of this measures use slightly different terms, including *discussed* and *mentioned* rather than *talked about*, as well as different response options, such as Never, Rarely, Sometimes, Frequently.
Comments	These measures provide a "deeper" measure of processing for labels and may be useful for comprehensive evaluations of labelling policies. Recommended, but not essential.
Construct	**(d) Thinking About Health Warnings**
Measure	"In the last month, how often have you thought about what the health warnings have to say?" (Never, Rarely, Sometimes, Often, All the time)
Sources	Canadian Cancer Society, 2001; Hammond *et al.,* 2003; Christie & Etter, 2004
Validity	Good face validity, convergent validity, and predictive validity for motivation to quit and future smoking behaviour when included as part of a composite measure.
Variations	"In the last month, have you ever thought about the warning labels or what they had to say when a cigarette pack wasn't in sight?" This variation of the measure requires a higher threshold of processing than the items above.
Comments	These measures provide a "deeper" measure of processing for labels and may be useful for comprehensive evaluations of labelling policies. Recommended, but not essential.

Table 5.28 Additional measures of labelling salience and processing

et al., 2003; Christie & Etter, 2004; Hammond *et al.*, 2004a; Health Canada, 2005; Koval *et al.*, 2005; Hammond *et al.,* 2006a; Hammond *et al.,* 2007a).

Measures of general "awareness" are typically endorsed by a vast majority of respondents, including non-smokers, regardless of the type of warning level. These questions are often used to examine the implementation, or "roll-out," of new package warnings following a change in policy. This information is critical for any population-based survey conducted shortly after the implementation of a new policy, given the uncertainty regarding when health warnings begin appearing on packages.

In contrast to general measures of awareness, the extent to which smokers notice, read, and think about the warnings appears to be highly dependent on the size, type, and location of the warning (Borland & Hill, 1997a; Health Canada, 2005; Hammond *et al.*, 2007a). These measures of processing are also subject to the implementation date. Several studies have used measures of processing collected from the same population over time and can be used to measure the "wear-out" (i.e. decrease in the salience of the warning labels) of health warnings (Health Canada, 2005; Hammond *et al.*, 2007a). Additional data of this type may help to answer the question as to whether the rate of decline among measures of salience is associated with design features, such as the size of warnings and the use of pictures. In most cases, these measures have been analyzed as individual items, although in one case a *depth of processing* scale was developed and tested (Hammond *et al.,* 2003). In that instance, nine items were used to create a scale to

Construct	Health Warnings - Eye Tracking
Measure	Participants wore eye-tracking equipment and viewed US cigarette advertisements with health warnings.
Sources	Fischer *et al.*,1989b; Krugman *et al.*, 1994
Validity	Good predictive validity for recall and recognition of health warnings
Variations	Viewing time serves as another measure of attention, where warnings are flashed on a screen and the amount of time is recorded (Peters *et al.*, 2007).
Comments	Eye tracking measures can help to identify the most salient design aspects of warning labels and serve as an objective measure of attention; however, these measures are limited to "laboratory" based research designs.

Table 5.29 Physiological Measures of Salience and Processing

measure cognitive processing labelled as "depth of processing." Responses to each of the nine items were rated using a 5-point Likert-type format going from "not at all/never" to "all the time/a lot" and values added to create an index. Examples of items included were "How carefully have you ever read the messages on the outside of a cigarette package?" and "How often have you thought about what messages on the inside of packages have to say?"

Although the wording of items is relatively similar across surveys, different time periods are used in both the question and the response option in many cases. For example, whereas some "noticing" questions refer to the past month, others refer to the past three months, or use no time reference at all (Tables 5.27 and 5.28). Nevertheless, findings from the same population are relatively similar across different question wordings (Canadian Cancer Society, 2001; Hammond *et al.*, 2004; Health Canada, 2005; Hammond *et al.*, 2007a).

Contents & emissions:

Several studies have assessed the extent to which smokers process emission information printed on the side of packages. These measures mirror the processing items used to gauge health warnings, although a more limited set has been used. Both studies of which we are aware, indicate that smokers are less likely to read or look at emission information than health warnings on the face of packages (Thompson *et al.*, 2006). More generally, it is unclear whether salience and processing type measures are as informative for emission labelling policies as for health warning policies. Unlike health warnings, which typically include a number of rotating health warnings, emission labelling is consistent across packages for a given brand. As a result, there may be little reason for smokers to read or attend to this information on a regular basis. As a consequence, we have not recommended a specific measure in this section.

Physiological measures of salience and processing:

Physiological measures have been used in conjunction with survey measures to quantify attention to and processing of health warnings. These measures have an advantage in that they are more "objective" given that they do not rely on self-reporting. In several cases, they have been used to compare the salience of warnings with package design or within the context of a tobacco advertisement. For

Construct	(a) Health Warnings - Location
Measure	Without looking at a cigarette package, where on the pack are the warnings or messages located?" (Open ended)
Source	Hammond *et al.*, 2003
Validity	Good face validity.
Variations	The same question has been asked without the prefix ("Without looking at a cigarette package…"), as well as with a diagram in self-completed surveys.
Comments	Useful measures for identifying basic knowledge about health warnings; however, it becomes complicated in jurisdictions with warnings on the inside and outside of packages. Unclear how emission and contents information should be treated, especially when provided by industry.
Construct	**(b) Health Warnings – Content**
Measure	"Without looking at a cigarette package, what specific health warning messages can you remember seeing on cigarette packages in Canada?" (Open ended)
Source	Health Canada, 2005
Validity	Good face validity.
Variations	The same question has been asked without the prefix, which is typically included in telephone surveys to ensure the participant is not looking at the package during the call.
Comments	Useful measures for identifying basic knowledge about health warnings and, potentially, for identifying individual messages that are particularly salient. However, this measure will be difficult to answer in jurisdictions with comprehensive health warnings, including multiple warnings on different areas of the package.

Table 5.30 Measures of Knowledge of Health Warnings

example, eye movements during exposure to an ad have been used as physiological indicators of attention to tobacco warnings. These measures that are directly linked to cognitive processing have been useful to investigate the relationship between visual attention and a more traditional communication measure (Krugman *et al.*, 1994) (Table 5.29).

Knowledge of health warnings

Items assessing smokers' knowledge of health warnings are among commonly used survey measures. Knowledge questions have been asked using unprompted recall (e.g. "Where are the warnings located?"), as well as using recognition tasks (e.g. "Please tell me which of the following warnings appear on cigarette packages…") (Table 5.30) (Hill, 1988; Richards *et al.*, 1989; Rootman *et al.*, 1995;

Construct	Emission Side Panel – Content
Measure	"Without looking at a cigarette package, can you name any chemicals or substances that are currently listed on cigarette packages in [country]?" (Open ended)
Source	Health Canada, 2003
Validity	Face validity.
Variations	A common alternative is to ask about the quantitative level of specific emissions, such as tar "Without looking at a pack, can you tell me the tar level of your cigarettes?"
Comments	This measure examines basic recall of emission information printed on packages, and is often compared against "objective" data collected from other sources in order to evaluate accuracy of self-report recall. This measure should be interpreted alongside measures on the comprehension and use of this information (described later).

Table 5.31 Measures of Knowledge of Constituents and Emissions

Construct	Health Warnings – Affective Reactions
Measure	"Have you experienced any fear as a result of the health warnings?" (Not at all / A little / A lot)
Source	Hammond *et al.*, 2004a
Validity	Good face validity; good predictive validity for future smoking behaviour.
Variations	Alternatives have used more comprehensive scales and asked about different affective reactions, including disgust and anger (Peters *et al.*, 2007).
Comments	Affective reactions have been evaluated to a greater extent in qualitative evaluations of warning labels; however, survey-based measures may be a key mediator of downstream measures of impact.

Table 5.32 Measures of Affective Reactions to Health Warnings

Borland & Hill, 1997a; Borland & Hill, 1997b; Robinson & Killen, 1997; Hammond *et al.*, 2003; Brown *et al.*, 2005; Health Canada, 2005; O'Hegarty *et al.,* 2006; Thompson *et al.*, 2006). Measures of unprompted recall for warning label content can be used to identify which individual warnings may be most effective. In jurisdictions with a large number of warnings, this task can be particularly helpful.

Many of these measures have been assessed among the general population, including among nonsmokers. Except for the few questions that refer to a respondent's "own" cigarette package, most measures of awareness and knowledge appear to work equally well among nonsmokers. Indeed, nonsmokers have been found to have

Construct	Health Warnings – Avoidance
Measure	"In the last month, have you made any effort to avoid looking at or thinking about the warning labels?" (Yes, No)
Sources	Hammond *et al*., 2004a; International Tobacco Control Policy Evaluation Survey (The ITC Project)
Validity	Good face validity; good predictive validity for future smoking behaviour.
Variations	Several follow-up questions may be asked of those who respond "yes" to the initial question, above. For example, "Have you made any effort to avoid the warnings by: (1) Covering the warnings up? (2) Keeping the pack out of sight? (3) Using a cigarette case or some other pack? (4) By not buying packs with particular labels?" (Yes, No to each question)
Comments	These measures can indicate the prevalence of avoidance behaviours and whether they reduce the effectiveness of warnings. The follow-up questions are only necessary for in-depth exploration of avoidance.

Table 5.33 Measures of Avoidance

surprisingly high levels of awareness and recall for prominent health warnings and picture-based warnings in particular (Health Canada, 2005). However, both recall and recognition of particular messages has been shown to be highly dependent on the complexity of the health warning and its implementation date. For example, virtually all Canadian smokers are aware of the health warnings on packages, although we are unaware of any research indicating that smokers have correctly been able to identify all 16 health warnings that appear on packages.

Analyses must take into account the consumption level when assessing knowledge of health warnings. Given the inevitable link between heaviness of smoking and viewing the warning labels, knowledge is likely to be greater among heavier smokers. This association is likely to be more pronounced within samples that include a broad range of smokers, and are likely to be greatest in studies that compare regular smokers with occasional or nonsmokers. The association between consumption and knowledge is also likely to be stronger in jurisdictions with a greater number and complexity of warnings. For example, packages in Canada carry information on the side panel, one of 16 health warnings on the outside of packages, and one of 16 additional warnings on the inside of packages. In such cases, a greater number of exposures will be required to recall various aspects of the warnings.

There are several limitations with measures of knowledge. First, when asking about the location of health warnings, one issue is whether respondents consider emission information, which may be printed on the sides of the package, as a health warning. Canadian data suggests that some smokers are aware of this information, but fail to cite it as a location. Second, in telephone or web-based surveys, some participants may have a pack visible as they respond to the survey. As a result, some measures explicitly ask smokers not to look at the package to avoid this situation to the extent possible. Third, measures of knowledge can often be difficult to compare across labelling policies. For example, smokers from the USA, where a total of four different text warnings appear on packages, have a much greater

Construct	(a) Health Warnings – Believability/Credibility
Measure	"Overall, do you believe the health warning message(s)?" (Not at all, A little, A lot)
Source	Health Canada Youth Smoking Survey (http://www.hc-sc.gc.ca/hl-vs/pubs/tobac-tabac/yss-etj-2002/index-eng.php)
Validity	Good face validity.
Variations	Other alternatives refer to the *accuracy, trustworthiness, credibility, believability* and *true/false* nature of the warnings or the *importance* of information (Cecil *et al.*, 1996; Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond *et al,* 2004a; Brown *et al.*, 2005; Health Canada, 2005; O'Hegarty *et al.*, 2006). Some surveys have also included more comprehensive, but also more time consuming, scales involving numerous items.
Comments	A useful, brief measure to examine credibility of message content. The measure can be used to examine whether different design and content features change the believability of information among smokers. This question can be asked of individual health messages, such as in qualitative or experimental research, or to refer to a set of warnings, as is common in population-based surveys. Note that responses to this item will also reflect denial, self-exempting beliefs, etc.
Construct	**(b) Health Warnings – Public Opinion/Support**
Measure	"Do you approve of the health warnings on cigarette packages?" (Yes, No)
Source	Borland & Hill, 1997a
Validity	Good face validity.
Variations	Other alternatives include measures of agreement with the warnings and references to *appropriateness or desire for more information* (Canadian Cancer Society, 2001; Hammond *et al.*, 2004a; Brown *et al.*, 2005., Health Canada, 2005)
Comments	This measure is a combination of previously administered questions and has yet to be administered exactly as worded. Though measures of public support or approval may be less important as a measure of effectiveness, they are a critical measure for regulators and policy makers, and for demonstrating support for more comprehensive policies.

Table 5.34 Measures of Credibility and Public Support

likelihood of correctly identifying all the messages than smokers in the United Kingdom where 16 different text messages appear on packages. The same issue arises in pre-post studies of a new labelling policy. For example, when Canada revised its labelling policy in 2000 to include pictures, the number of individual messages doubled from eight to 16 (not counting 16 additional messages that appeared on the inside of packages). In such cases, neither the total number nor the proportion of messages correctly identified, provide a suitable basis for comparing policies given that the denominator is different. Moreover, it is both time consuming

Construct	Health Warnings – Thinking About Health Risks
Measure	"To what extent, if at all, do the warning labels make you think about the health risks of smoking?" (Not at all, A little, A lot)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; good convergent validity; associated with strength of policy.
Variations	Similar questions ask about the extent to which warnings affect the level of concern or *worry* about health risks.
Comments	A key mediator of the effectiveness of health warnings. This should be considered among the essential measures.

Table 5.35 Measures of Health Knowledge and Perceived Risk

Construct	Emissions Information – Comprehension
Measure	"If you were to look for a safer or less harmful cigarette, would you use information about the amounts of chemicals listed on the cigarette packs to help you find a less harmful brand?" (Yes, Maybe, No)
Sources	Gori, 1990; Health Canada, 2003
Validity	Good face validity.
Variations	Similar questions ask smokers to compare different tar levels of cigarettes in terms of delivery and health risks.
Comments	A critical measure to evaluate emission policies that include quantitative emission levels. The question can also be used to refer to specific emissions, such as tar or nicotine. This measure is essential in any survey that also asks about recall or awareness of emission numbers on packages. The current wording can be used to refer both to descriptive (i.e. text-based) and quantitative emission information.

Table 5.36 Measurement of Comprehension of Emissions Information

and awkward to prompt survey respondents for 16 different warnings.

Finally, some knowledge measures may not work across all survey modalities. For example, Krugman and Robinson presented participants with diagrams of various warnings in a recognition task (Krugman *et al.*, 1994; Robinson & Killen, 1997). Any such measures, which require visual information to be presented to participants, must be administered either face-to-face or using web-based modalities.

Construct	(a) Light / Mild Descriptors – Comparative Risk
Measure	"Light cigarettes are less harmful than regular cigarettes." (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree)
Source	The ITC Project
Validity	Good face validity; good convergent validity (Borland *et al.*, 2004).
Variations	This question can be adapted to refer to other descriptors, such as *mild* or *smooth*. In some cases, the terms *light* and *mild* are used in the same question. Alternatives ask smokers about differences in the "tar" or "nicotine" of *light* versus *regular* cigarettes (Smokers of light cigarettes take in less tar than smokers of regular cigarettes). These measures have been widely used, but require a basic familiarity with tar and nicotine, which may not exist in all smokers in some jurisdictions (Kozlowski *et al.*, 1998b; Shiffman *et al.*, 2001; Hamilton *et al.*, 2004). Other alternatives have asked smokers to report the number of *light* cigarettes that would need to be smoked to equal the harm from 10 regular cigarettes; however, this approach requires a level of numerical literacy beyond the capacity of smokers in many jurisdictions (Kozlowski *et al.*, 2000; Shiffman *et al.,* 2001).
Comments:	This is an essential construct, although there is no single "gold standard" question for its measurement. The recommended measure has been selected because is it the most direct and may be most appropriate for smokers in low- and middle-income countries. Nevertheless, the question may need to be preceded by a general awareness questions (e.g. "Have you ever heard of *light* cigarettes?") in some markets or rural areas. There are also issues with the interpretation of this measure in jurisdictions where *light* and *mild* descriptors have been prohibited.
Construct	**(b) Light/Mild Descriptors – Addiction**
Measure	"Light cigarettes are less addictive than regular cigarettes." (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree)
Source	The ITC Project
Validity	Good face validity; good convergent validity (Borland *et al.*, 2004).
Variations	This question can be adapted to refer to other descriptors, such as mild or smooth. In some cases, the terms *light* and *mild* are used in the same question.
Comments	A straightforward question with the same format and response options as above. A recommended question to address perceptions of *light* and *mild* cigarettes, but not as essential as the comparative risk question, above.

Table 5.37 Measures of Light, Mild, and Brand Descriptors

Construct	Brand Appeal – Health Warnings
Measure	"Do you think the new warnings make cigarettes packages look less attractive, more attractive, or has it made no difference to their attractiveness?" (Not at all, A little, A lot) "How often have you put your cigarette package away because you didn't want others to see the warning on the package? Have you done this?" (Never, Sometimes, Often)
Source	Canadian Cancer Society, 2001
Validity	Face validity.
Variations	Alternatives refer to quality of advertisements with and without warnings, whether youth would want to "use" the product, intentions to purchase the product in the future, and a measure of perceived economic values of brands (Hyland & Birrell, 1979; Brubaker & Mitby, 1990; Canadian Cancer Society, 2001; Willemsen *et al.*, 2002; Thrasher *et al.*, 2007). "Attractiveness" scales have also been used (Loken & Howard-Pitney, 1988).
Comments	These measures provide a straightforward evaluation of whether health warnings have altered the general appeal of packaging. The second of the two measures has a higher threshold and represents a more distal measure of appeal, which may also tap into social norms. Both of the measures are recommended for surveys that wish to provide a comprehensive evaluation of warnings, but are not essential.

Table 5.38 Measures of Brand Appeal

Constituents & emissions:

A number of studies have examined whether smokers can recall the emission information commonly printed on the side panel of cigarette packages (Table 5.31) (Chapman, 1986; Cohen, 1996b; Health Canada, 2003; O'Connor *et al.*, 2006c). These items typically ask participants to name the emissions printed on packages using unprompted recall tasks, or ask them to report the number associated with a particular emission (usually "tar"). The data indicates that many smokers have a general awareness that tar and nicotine numbers may be printed on the package, but few are able to recall the tar or nicotine levels printed on their usual brand of cigarettes. To our knowledge, no measures have been developed to measure smokers' knowledge of tobacco contents.

Affective reactions to health warnings:

Research in the field of health communication indicates that messages with emotionally arousing content are more likely to be noticed and processed by smokers (Witte & Allen, 2000). Strong emotional responses to messages are also associated with greater behaviour change when supportive or "efficacy" related information is also presented. To date, several studies have used measures of affective reactions to assess the impact of warnings labels (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; Environics Research Group, 2003; Hammond *et al.,* 2004a; Health Canada, 2006; Peters *et al.*, 2007). These measures are common in qualitative evaluations of individual warning labels and have been particularly influential in development of picture-based warnings in

Construct	(a) Changes in Foregoing – Health Warnings
Measure	"In the last month, have the warning labels stopped you from having a cigarette when you were about to smoke one?" (Never, Once, A few times, Many times)
Sources	Borland & Hill, 1997a; Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity; associated with strength of policy.
Variations	Similar measures have referred to not smoking when *tempted.*
Comments	This question has a lower "threshold" than other measures that assess the behavioural effects of health warnings.
Construct	**(b) Reductions in Smoking – Health Warnings**
Measure	"Are you smoking any less or more as a result of the new warnings, or are you still smoking the same amount?" (Less, Same amount, No difference)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity; associated with strength of policy.
Variations	Similar measures have referred to not smoking when *tempted.*
Comments	The wording "as a result of the warnings" needs to be emphasized when asking this question. This item is not intended to provide a precise measure of changes in consumption as a result of the warnings; changes in consumption happen in response to a wide range of related factors. However, this question does provide a good general measure of the extent to which smokers have been affected by the warnings.
Construct	**(c) Likelihood/Motivations to Quit**
Measure	"To what extent, if at all, do the warning labels on cigarette packs make you more likely to quit smoking?" (Not at all, A little, A lot)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity.
Variations	Alternatives refer to *motivations to quit* and *thinking about quitting*, with some differences between response categories.
Comments	The recommended wording refers directly to the likelihood of quitting smoking, which is somewhat broader than motivation alone. In practice, however, there appears to be few differences with regards to how these measures perform in practice given the consistency of findings from similar samples. The question has the potential to provide a very good summary measure of the self-reported impact of health warnings and should be considered within the core set of items to evaluate labelling policy.

Table 5.39 Measures of Behavioural Outcomes

Construct	(d) Quit Attempts & Abstinence
Measure	"To what extent, if at all, were the following reasons for your current quit attempt…warning labels on cigarette packages?" (Not at all, A little, A lot)
Source	The ITC Project
Validity	Good face and convergent validity.
Variations	Alternatives have also asked about the effect of the warnings on staying quit in the future. This question can be asked as part of a list of reasons for quitting, which provides some useful context on the relative influence of other potential influences on quitting.
Comments	Retrospective measures, such as this, should be interpreted with caution given that they are subject to recall biases, particularly as the time since the quit date increases. In addition, smokers often cite a number of complementary reasons for quitting and endorsement of this item does not mean that the quit attempt is solely attributable to health warnings.

Table 5.39 Measures of Behavioural Outcomes

several jurisdictions. Measures of negative emotions, including fear and disgust, have also been used in population-based surveys and shown to predict future cessation-related behaviour (Table 5.32). Overall, measures of emotion have considerable promise as a proximal measure of effectiveness which can be used in both qualitative and quantitative research.

Avoidance:

Warnings that result in unpleasant emotions may lead some smokers to avoid the warnings. Indeed, several studies indicate that a considerable portion of smokers make some attempt to avoid the warnings, including covering or hiding the warnings, using another case, or requesting different packs to avoid particular warnings. In some jurisdictions, tobacco manufacturers have been accused of marketing covers specifically intended to cover picture-based warnings, prompting calls for regulatory bans on the sale of such covers (Table 5.33) (Wilson *et al.,* 2006).

Although avoidance behaviours may be undesirable to some extent, these examples of fear control behaviour do not necessarily reflect an adverse outcome or inherent weakness of package warnings. Research has demonstrated that avoidant behaviours and attempts at thought suppression often have the opposite effect of increasing the presence of the unwanted thoughts (Wegner, 1994). In the context of the warning labels, avoidant behaviour might be more reasonably interpreted as a measure of effectiveness. Indeed, if the warnings were ineffective in communicating the threatening consequences of smoking there would be no reason to avoid them. Furthermore, one study found that smokers who attempted to avoid the warnings were no less likely to see the warnings, think about them, or engage in cessation behaviour at a 3-month follow-up (Hammond *et al.*, 2004a).

Credibility & public support:

In order to be effective, the health information presented in warnings must be credible. The credibility of warnings relates not only to the health information contained in a warning, but also to its design and source or attribution. Some have even speculated that there may be a trade-off between the vividness of the information in health warnings and its credibility among smokers. In others words, if

pictures and text become too striking or graphic, smokers may begin to question the accuracy of the information and become more resistant to the messages.

Although some validated scales have been used to evaluate the believability of health warnings (e.g. Beltramini, 1988; Loken & Howard-Pitney, 1988; Cecil *et al.*, 1996), many studies have used single questions with face validity (Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond *et al.*, 2004a; Brown *et al.,* 2005; Health Canada, 2005; O'Hegarty et al., 2006; Peters *et al.*, 2007). Together, the findings suggest that health warnings represent a credible source of information, particularly when attributed to a well-respected department of health, or a well-respected non-governmental authority, such as a cancer society (Guttman & Peleg, 2003; Health Canada, 2003; BRC Marketing & Social Research, 2004). The levels of credibility do not appear to be associated with the type or design of warning labels; just like for text-based warnings, smokers report high levels of believability for graphic picture-based warnings as well.

Several studies have also sought to assess general measures of public support for health warnings (Borland & Hill, 1997b; Brown *et al.*, 2005; Hammond *et al.,* 2004a; O'Hegarty *et al.*, 2006). To our knowledge, two items have been developed to examine support among smokers for emission labelling ("Overall, do you believe the health warning message(s)?" and "Do you approve of the health warnings on cigarette packages?") (Health Canada, 2001; Health Canada, 2003). Public opinion data may be particularly effective for policy makers in gauging political support for new or existing labelling polices (Table 5.34).

Health knowledge & perceived risk

The primary objective of cigarette warning labels is to communicate the health effects from smoking. Thus, measures of health knowledge and perceived risk represent critical components in any evaluation of health warnings (Table 5.35). To date, studies have taken two main approaches to measuring the impact of warnings on health knowledge. One approach is to ask participants to self-report whether health warnings have changed the extent or frequency with which they think or worry about the health effects of smoking. Alternatively, some studies have assessed health knowledge directly and examined changes over time or across jurisdictions in levels of knowledge. Given the number of health effects caused by smoking, we are unaware of any study that has attempted to measure a complete list. However, studies typically measure beliefs about a range of specific health effects to determine knowledge levels. Some studies have included "bogus" health effects in the list in order to identify response bias. Most lists include "major" health effects, such as lung cancer and heart disease, as well as health effects on nonsmokers, and lesser-known health effects. Including lesser-known health effects can be particularly effective in attributing changes in knowledge to specific labelling policies. Ideally, longitudinal studies, assessing changes in health knowledge, would also select the health effects based upon the effects that are targeted in the warnings. In other words, studies should include health effects that: a) are already included on packages at baseline (before policy change) and will remain on packages at follow-up; b) health effects that are not on packages at baseline, but will appear at follow-up; and c) health effects that are not on packages at either baseline or follow-up. This type of design provides a measure of specificity with respect to changes in labelling policies.

A similar approach has been taken with respect to emission information. At least one study has examined whether knowledge of the emissions in tobacco smoke is higher in jurisdictions where they are printed on the package (Hammond *et al.*, 2006a). As with health effects, lists should include emissions that are, and are not, printed on packages, in order to examine the specificity of the effect.

Overall, research conducted to date suggests that increases in the size, number, and content of warnings are associated with greater thoughts about the health risks of smoking (Health Canada, 2005; Hammond *et al.*, 2007a).

More prominent warnings have also been associated with increased knowledge for specific health effects (Borland & Hill,1997a; Hammond, 2006a). Most of these findings derive from population-based surveys, although one study reported significantly higher beliefs about health effects following presentation of graphic versus text warnings within an experimental setting (O'Hegarty *et al.*, 2006).

Constituents & emissions:

A number of studies have sought to examine the extent to which smokers understand and interpret quantitative cigarette emission information (Table 5.36). These studies ask smokers to report either the "meaning" of the numbers, or the extent to which the numbers translate into differences in exposure from different brands (Gori, 1990; Cohen, 1996a; Health Canada, 2003; Thompson *et al.*, 2006). Other questions ask smokers to predict the health consequences of different tar levels, without explicit reference to labelling policies (Gori, 1990; Cohen, 1996a). Indeed, a number of studies on this topic were conducted in the USA, where there are no mandatory requirements to print emission levels on packages, they appear on packages less than 15% of the time, and are at the discretion of the manufacturer (Davis *et al.,* 1990).

Regardless of the jurisdiction or the labelling policy, the findings indicate that smokers have very little or no understanding of the meaning of the emission levels, although a substantial proportion associate health benefits with lower numbers. This type of data is critical to place measures of knowledge into context; prominent labelling that succeeds in increasing knowledge of emission levels is of little value if smokers do not understand the meaning of these numbers. Indeed, the data appear to indicate that communicating quantitative emission levels promotes erroneous perceptions about exposure levels and health risks that can be expected from different products. In general, this set of findings underscores the importance of assessing more than basic recall of information (Figure 5.30).

Light & mild descriptors:

A variety of surveys have examined perceptions of "light" and "mild" brand descriptors (Kozlowski *et al.*, 1998b; Kozlowski *et al.*, 2000; Ashley *et al.*, 2001; Shiffman *et al.*, 2001; Etter *et al.*, 2003c; Borland *et al.*, 2004; Hamilton *et al.*, 2004). Both quantitative and descriptive measures have been used to assess the health consequences of smoking "light/mild" cigarettes. Several studies have asked smokers how many light or ultra-light cigarettes would need to be smoked to inhale the equivalent level of tar as regular cigarettes. Some of these measures used "10 cigarettes" as a reference point, whereas others were open-ended. Smokers have also been asked to make comparisons between "light/-ultra-light" and "regular" brands using qualitative or descriptive categories to describe exposure levels and health risks. These qualitative response categories have also been used to compare perceived sensory properties and addiction levels of "light/mild" cigarettes compared to "regular" brands. At least one study combined items to create a "sensory" index and a "health effects" index (Shiffman *et al.*, 2001). Overall, both qualitative and quantitative measures appear to yield similar findings, and indicate that a substantial proportion of smokers perceive health benefits from cigarettes with "light" and "mild" descriptors (Table 5.37).

At least one study, the International Tobacco Control Policy Evaluation Survey (the ITC Project) (Borland *et al.*, 2004), has adopted an alternative approach to comparative estimates. Rather than asking smokers to compare "regular" and "light" cigarettes, participants were asked to compare their "usual" brand with regular cigarettes (e.g. "Do you think that the brand you usually smoke, [current brand], might be a little less harmful, no different, or a little more harmful, compared to other cigarette brands?"). Separate items were used to collect the name, descriptors, and relevant attributes of participants' "usual" brand. This approach has the benefit of personalizing the question, and is particularly useful to implement following the removal of "light" and "mild" terms, at which point questions with direct reference to "light" and

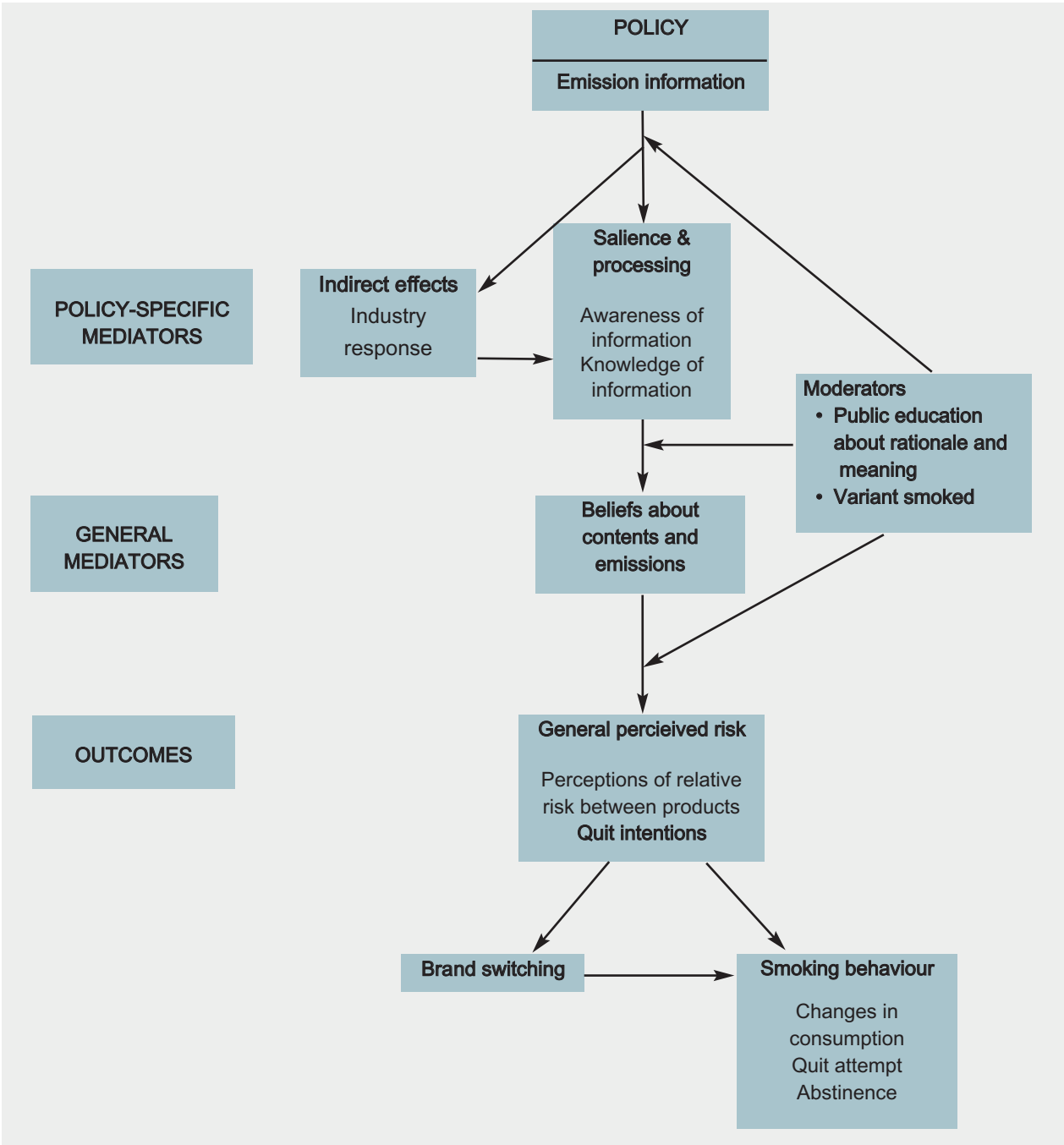


Figure 5.30 Conceptual Framework for the Evaluation of Emissions and Contents Labelling Policies

"mild" cigarettes become awkward and confusing. The question also has a broader frame of reference. This is an advantage in the sense that it captures the effect of other potential misleading descriptors or product elements. The disadvantage is that information on the respondents' own brand must also be available (see Section 5.4), and there is less specificity with respect to the brand elements that underlie differences in perceptions of risk. A similar conceptual approach has recently been taken with respect to evaluating print advertisements. Rather than asking smokers to compare the risks implied by the expressions "light" versus "regular" cigarettes, respondents were asked to rate the perceived risk to their health derived from advertisements for different products, and the ratings for advertisements of "light" versus "regular" cigarettes were compared (Hamilton *et al*., 2004). In most cases, follow-up questions may be necessary to identify which specific elements underlie perceptions of reduced harm.

Descriptors other than "light/-mild" are likely to receive increased attention in the coming years, particularly within jurisdictions where "light/mild" terms have already been prohibited. To our knowledge, only one study has developed measures to evaluate health perceptions based on other brand descriptors, including the words "smooth" and "ultra" (Thompson *et al*., 2006). Furthermore, studies with a focus upon brand descriptors in jurisdictions that have banned "light" and "mild," may wish to consider additional measures that examine the substitution of terms in their place. Market-based research, such as cataloguing the information printed on packages, can provide "objective" data on the substitution of terms which may be helpful in interpreting self-reported brand data (see Section 5.4).

Largely, the selection of measures in this area may depend upon the current state of policy more so than other areas (Figure 5.31).

Brand appeal

Health warnings target psychosocial variables other than perceived risk and health knowledge. More recent labelling policies include themes of addiction, industry manipulation, aesthetic costs, financial costs, and cessation beliefs, among others. A range of psychosocial measures have been developed to assess each of these constructs, although these measures have rarely been used to evaluate warning labels.

One area that has been explored is the impact of health warnings on measures of brand appeal (Table 5.38). In theory, replacing brand imagery with health warnings has the potential to change perceptions of the cigarettes and packaging. To date, the limited findings in this area appear to support this hypothesis, although it has yet to be explored in much depth with respect to warnings on packages (Hyland & Birrell, 1979; Loken & Howard-Pitney, 1988; Brubaker & Mitby, 1990; Hammond *et al*., 2004b; Thrasher *et al*., 2007). Future research might also explore whether larger graphic health warnings undermine the visual appeal of cigarette displays at retail outlets.

Behavioural outcomes

There are several approaches to predicting "downstream" cessation-related outcomes from health models. As with health effects, some studies have used measures of processing and knowledge of the warnings, and modelled their effects on motivation to quit and patterns of smoking behaviour (see Section 3.1 for measures of tobacco use and Section 3.2 for psychosocial outcomes). This has produced significant findings in longitudinal studies to date (Hammond *et al.,* 2003). However, this approach is somewhat limited when it comes to evaluating changes in health warnings. Unless both survey waves are conducted when the same set of health warnings is on the pack, the baseline measures of processing or knowledge relate to the "old" warnings, whereas any cessation-related activity at follow-up presumably reflects the impact of the "new" warnings.

An alternate strategy that can also be used in cross-sectional studies is to ask smokers to directly report the extent to which warnings have influenced their motivation to quit and smoking behaviour (Borland & Hill, 1997a;

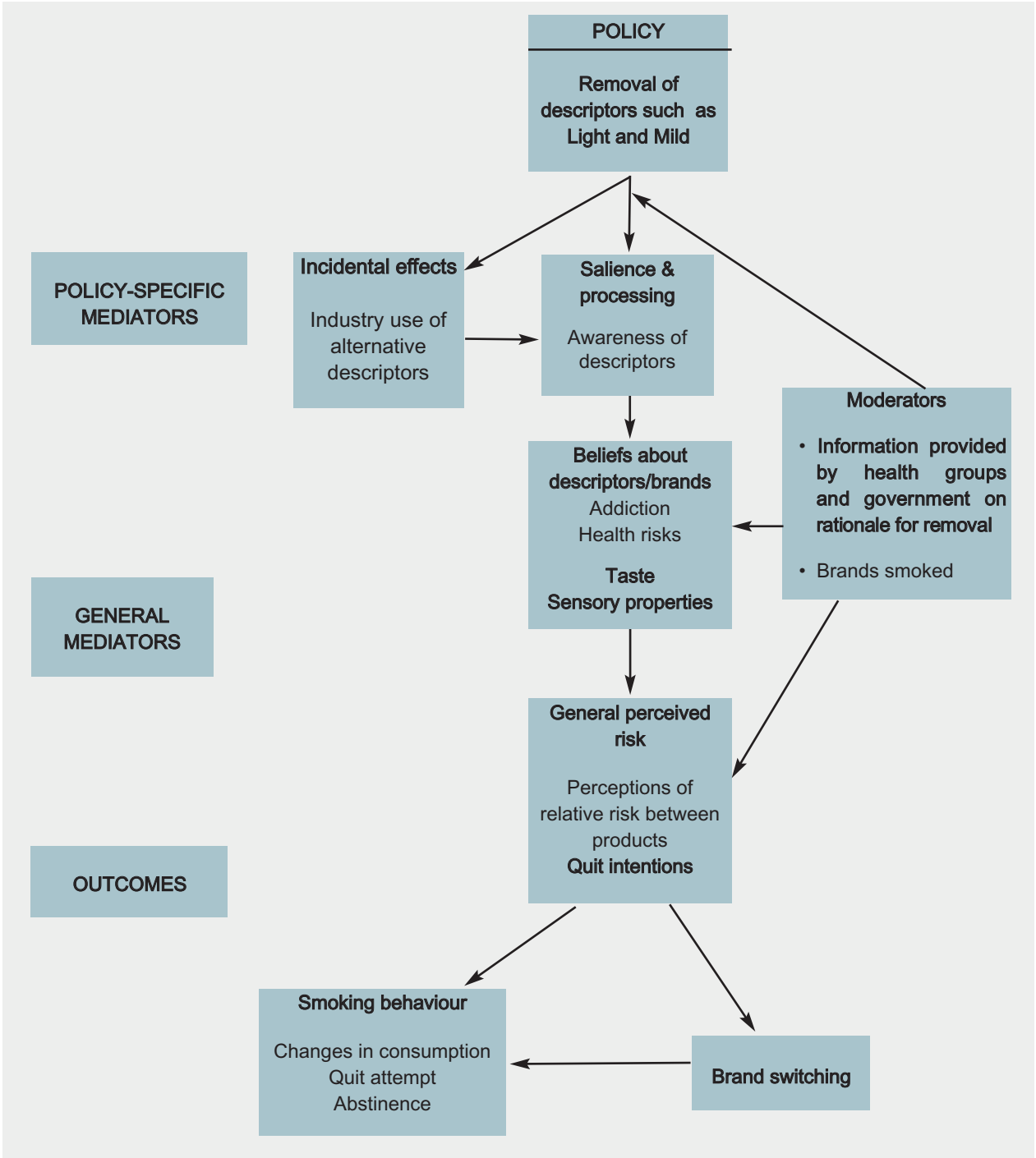


Figure 5.31 Conceptual Framework for the Evaluation of Brand Descriptor Policies

Canadian Cancer Society, 2001; Health Canada, 2005; Koval *et al*., 2005; Willemsen, 2005; O'Hegarty *et al*., 2006). This approach does not have the same validity in terms of measuring actual changes in smoking behaviour, although it can be used to examine changes across labelling policies.

A third alternative is to examine changes in prevalence rates, or population-based cessation activity, before and after the implementation of new warnings. To our knowledge, this approach has been used in only one study to date: Gospodinov & Irvine (2004) reported no discernable changes in prevalence rates, and a reduction of two cigarettes per week among smokers in the months following the implementation of pictorial health warnings in Canada. However, as described earlier in this section, there are serious problems in attributing changes in national level trends to changes in health warnings, or any other individual policy measure. Indeed, as Gospodinov & Irvine note, there were significant changes in price over the same period of time, as well as considerable sub-national tobacco control activity over the same time period.

Yet another approach to measuring the impact of warnings on cessation behaviour has been to look at changes in the use of cessation services as they relate to information on warnings labels. Research conducted in the UK and the Netherlands has examined changes in the usage of national telephone helplines after the contact information was included in package health warnings. Each of these studies reports significant increases in call volumes (Willemsen *et al.*, 2002; Department of Health, 2006).

Finally, several items have been created for use among former-smokers. Typically, these items ask about various reasons for quitting, including whether the health warnings either motivated them to quit or have helped them to remain abstinent (Canadian Cancer Society, 2001; Hammond *et al*., 2004b; O'Hegarty *et al*., 2006; Thompson *et al*., 2006). These measures are, however, subject to recall bias and should be interpreted with particular caution (Table 5.39).

Formative research

Formative research is often conducted to help identify the content and design of new health warning policies. Regulators must decide what health effects to communicate, how many, and how to present this information to smokers on the package. Although population-based surveys may help to guide these decisions, qualitative research is typically undertaken as part of the policy development process.

The most common approach has been to conduct a series of focus groups (i.e. semi-structured interviews conducted within a group setting). Focus groups have two important advantages over population-based surveys: 1) participants can be presented with visual stimuli, including examples of health warnings in a way that is not possible with telephone based surveys; and 2) focus groups are well suited to open-ended questions and allow for more in-depth discussion than structured surveys. In many cases, focus groups are also used as a way to evaluate the effectiveness of health warnings on sub-groups, including younger smokers and those from lower socio-economic groups. The primary disadvantage of focus groups is that the findings can be hard to summarize in a systematic fashion, which complicates comparisons across groups and settings. As a result, conventional validity tests for quantitative data can not be conducted with focus group findings. Nevertheless, qualitative findings help to complement quantitative research in this area, and represent an important step in the development of new labelling policies.

Qualitative research has examined many of the same themes as population-based surveys, and other quantitative methods. These include general knowledge of the warnings, such as the content and location, the emotional impact of warnings, as well as their general salience and noticability (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; CRÉATEC, 2003; BRC Marketing & Social Research, 2004; Health Canada, 2006). In many cases, these studies have presented different health warnings to participants in order to make direct comparisons between labelling policies. These designs have proven particularly effective at comparing the emotional reactions

elicited by picture versus text warnings, for example (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; BRC Marketing & Social Research, 2004). Focus groups have also provided critical information regarding the meaning and comprehension of the information communicated in labelling policies. For example, focus group measures developed, on behalf of Health Canada, have helped to demonstrate that, even though most Canadian smokers are aware of emission information on the side of packages, very few understand the actual meaning of the information (Environics Research Group, 2003). Indeed, judging by the findings of the focus group, most Canadian smokers are misusing the emission information. Thus, carefully constructed focus group measures can provide "deeper," more comprehensive measures of meaning that are difficult to ascertain through structured population-based surveys.

Industry documents

Internal tobacco industry documents represent a potentially rich source of information about the effectiveness of tobacco control policies. There are several informative reviews of industry activities and documents on product labelling, including many related to brand descriptors such as "light" and "mild" (Slade, 1997; Pollay, 2001; Pollay & Dewhirst, 2002; Wakefield *et al.*, 2002; Chapman & Carter, 2003; Alechnowicz & Chapman, 2004). However, to date, no comprehensive review of packaging issues related to labelling policies has been undertaken.

Summary

Few of the measures used to evaluate warning label policies have undergone formal psychometric analyses. Much of the literature in this area has been conducted on behalf of regulators, which may account for the lack of "formal" tests of validation more common to academic research. In addition, different studies have used different measures to assess the same construct. In many cases, measures differ in the wording of questions and in the time references used in measures, such as noticing and awareness. This complicates comparisons across surveys and across labelling policies. However, most measures have high face validity and several have shown good predictive validity for downstream outcomes, including knowledge of health effects and self-reported motivation to quit, and cessation behaviours. In addition, the consistency of the findings across studies and survey modalities suggests that the differences in the measures have only a modest effect on outcomes of interest. Nevertheless, virtually all of the constructs would benefit from further developmental work, including the standardization of the wordings across surveys.

Implications for study design & analysis:

No single study research design is adequate to evaluate the impact of labelling policies. Given the challenges inherent in evaluating national level policies, individual studies are inevitably subject to a range of limitations. However, when taken collectively, the range of designs constitute a persuasive body of evidence demonstrating the effectiveness of comprehensive health warnings. Qualitative methods, including focus groups, are essential for informing the early stages of design and generating new insights into labelling policies. Experimental research is best suited to drawing direct comparisons across warnings and to isolating the effectiveness of individual design and content features. For this reason, experimental research provides the highest level of internal validity. Alternatively, population-based surveys have the highest external validity and may provide the most comprehensive measures of effectiveness given adequate designs. External validity is particularly important in the case of warning labels, which operate over repeated exposures that are tied to smoking behaviour. The pattern of exposure is the defining feature of product warnings and one that is impossible to replicate in a "laboratory" environment. As a result, the central question of whether labelling policies influence beliefs, attitudes, and behavioural change can only be

assessed with population-based surveys. The inferences that can be made from these surveys are considerably enhanced within longitudinal and quasi-experimental designs, as discussed in Section 2.1.

Priorities for future work:

As countries begin to implement restrictions on misleading packaging elements, research must begin to examine elements other than "light" and "mild" brand descriptors. These include other potentially misleading elements, such as the use of colour-coding and package designs that falsely convey differences in strength. To date, very limited work has been conducted outside the tobacco industry on these issues. There is an immediate need to develop measures that can examine these issues within population-based samples, especially within jurisdictions where "light" and "mild" descriptors have already been prohibited.

A second priority for future research is to examine contents and emission information more closely. Up to now, much of the existing research has focussed upon awareness and understanding of ISO tar and nicotine numbers. There is an urgent need for measures to evaluate new approaches to communicating contents and emission information. Population-based studies should be conducted within jurisdictions that have developed novel policies, such as communicating emission information using descriptive, rather than quantitative means. Greater experimental and qualitative work must also be undertaken to explore how smokers interpret and use this information, and to compare different approaches more systematically. These issues are directly relevant to the ongoing debate regarding how to communicate the risks of combustible versus non-combustible tobacco products. Historically, emission information has been used by smokers to evaluate the relative risks of different products. As emission and content labelling policies are developed for the full range of tobacco products, regulators will need to consider the delicate issue of what fundamental message they wish to communicate to smokers. Quantitative emission and content information will inevitably be interpreted as indicators of risks, unlike descriptive information that is uniform across products.

In addition to developing new survey measures, existing measures must be administered more widely, as a greater number of countries prepare to implement the provisions within Article 11 of the FCTC. In particular, few of the measures reviewed in this section have been assessed among smokers in low- and middle-income countries.

Finally, measures should be developed to examine the impact of the cessation information that is included in many labelling policies. Cigarette packages are among the most prominent vehicles for disseminating cessation services and efficacy-related information. These measures may include survey based measures, as well as indicators from other data sources, such as usage rates from telephone quitlines or web-based services.

Recommendations

Comprehensive evaluations of health warning labels should include recommended items from each of the key constructs (see above). Population-based surveys, seeking a more limited evaluation of health warnings, should include proximal measures of noticing, along with intermediate measures of perceived risk or health knowledge. Although measures of general awareness and knowledge of health warnings can be informative, these measures should be used with caution for the purpose of comparing labelling policies.

Evaluations of brand descriptors, and other packaging elements, should represent a priority for tobacco control policy. In addition to examining "light" and "mild" descriptors, research should consider other potentially misleading terms, as well as brand elements such as colour and package design. Unlike health warnings, these policies require the removal of information from the package and present challenges in the wording of survey measures. There is an immediate need to develop measures that can address these issues as more countries implement recommendations under Article 11 to

prohibit misleading package elements.

Policies to communicate emissions and content information via packages, also present unique evaluation challenges. Unlike health warnings, measures of salience and processing for this type of information are of limited value. Rather, evaluations should focus upon the meaning and use of emission and content information. Given the lack of research in this area, and the lack of consensus regarding the best policy approach, there is a particular need for formative research in this area.

Overall, the selection of measures to evaluate tobacco labelling policies will depend upon the method and scope of the evaluation, as well as the specific policy context.

5.6 Measures to assess the impact of anti-tobacco public communication campaigns

Introduction

Public communication campaigns are used to improve awareness, knowledge, and understanding of an issue, in an attempt to influence individual behaviour, build support for, and contribute to policy and social change. Carefully monitoring the implementation and outcomes of campaigns is essential to ensuring their effectiveness and demonstrating their contribution to a specific public health outcome. This section summarizes the main components of individual behaviour change and public will campaigns, briefly describes the theory and practice of public communication campaigns and their evaluation, and provides approaches for evaluating each component to determine impact, from planning and development through implementation and demonstrating results. Specific measures are identified for use as indicators of the achievement of proximal and intermediate outcomes of public communication campaigns. However, the key to measuring the impact of public communication campaigns is articulating clearly at the outset what the campaign is intended to accomplish, who the campaign is intended to reach, what the campaign is intended to cause, and what communication and evaluation strategies will be used.

A comprehensive public communication campaign will include multiple components and demand extensive resources, particularly for media production and placement (Atkin, 2001; Coffman, 2002; Dorfman *et al.*, 2002). These components may include resources for advertisement production and placement across a range of media; development and use of press materials and press events; advocacy activity to influence how messages are framed and interpreted; and community action to make messages locally relevant, compelling, and supportive of campaign goals. However, specific campaign components may be implemented independently and, depending on the desired outcomes, may be nearly as effective as a comprehensive campaign. Depending on the aims of the campaign, and the resources and opportunities of the local jurisdiction (nation, province, state or community), specific components or combinations of components will be more relevant. This section provides guidance on evaluation methods for use in planning and implementing a public communication campaign in order to increase the likelihood of success. It will also serve as guidance on measures to be used to demonstrate the effectiveness of the campaigns in achieving more proximal outcomes associated with the WHO FCTC Article 12 directives (WHO, 2003; Figure 5.32).

Selecting measures of effectiveness and demonstrating them are easiest when a campaign is grounded in a change theory that describes a logical progression from activities to outcomes. Measures of effectiveness then can be selected to coincide with specific expected outcomes, as described in Figures 5.33 and 5.34. For example, a public communication campaign designed to increase support for and promote the enactment and effective implementation of a smoke-free air law might include:

1. Television, radio, and print advertising about the health hazards associated with exposure to tobacco smoke, with measures of effectiveness demonstrating that the target audience saw or heard and understood the message and assimilated the information (i.e., awareness, attitudes, beliefs, or knowledge increased or were reinforced).
2. Contacts with news, health, community reporters, and editorial staff to encourage news, editorial, and community interest

Each Party shall promote and strengthen public awareness of tobacco control issues, using all available communication tools, as appropriate. Towards this end, each Party shall adopt and implement effective legislative, executive, administrative or other measures to promote:

(a) broad access to effective and comprehensive educational and public awareness programmes on the health risks including the addictive characteristics of tobacco consumption and exposure to tobacco smoke;
(b) public awareness about the health risks of tobacco consumption and exposure to tobacco smoke, and about the benefits of the cessation of tobacco use and tobacco-free lifestyles as specified in Article 14.2;
(c) public access, in accordance with national law, to a wide range of information on the tobacco industry as relevant to the objective of this Convention;
(d) effective and appropriate training or sensitization and awareness programmes on tobacco control addressed to persons such as health workers, community workers, social workers, media professionals, educators, decision-makers, administrators and other concerned persons;
(e) awareness and participation of public and private agencies and nongovernmental organizations not affiliated with the tobacco industry in developing and implementing intersectoral programmes and strategies for tobacco control; and
(f) public awareness of and access to information regarding the adverse health, economic, and environmental consequences of tobacco production and consumption.

WHO (2003)

Figure 5.32 WHO FCTC Article 12: *Education, communication, training and public awareness*

stories about the dangers of tobacco smoke and conveying support for smoke-free policies. Proximal outcomes might be the news and special interest stories and editorials that are printed or aired addressing the policy goals.

3. Media advocates might use similar public relations strategies focused on media outlets in particular legislative districts that are known to be accessed by influential leaders. Outcomes might be documents from records of public comments by the targeted decision makers.
4. Community groups and members may be organised to host community education events, meet with political representatives, offer personal testimonials of the value of smoke-free air policies or adverse impacts of tobacco smoke exposure. Proximal outcomes of these strategies could include media coverage of community events, opinion polling, intercept interviews, or other indicators of community attitudes, and meetings with or other engagement of local decision makers.

Together, these coordinated actions, promulgating a clear and consistent message and demand for policy action, constitute a comprehensive public communication campaign to advance the public health as outlined in the WHO FCTC; specifically, as directed in Article 12 (Figure 5.32). This section describes the use of public communication campaigns to advance these Article 12 directives and measures of whether the campaign has contributed to specific goals.

Components of a public communication campaign

Public communication campaigns tend to be divided into two types, each emphasizing somewhat different strategies and outcomes: individual behaviour change campaigns, and public will or public engagement campaigns (Coffman, 2002). Individual behaviour change campaigns seek to change the types of behaviours that lead to personal or social problems or instill behaviours that will improve individual or social well-being (Coffman, 2002). Public will campaigns, on the other hand, focus on motivating public officials to take policy action, which in turn will motivate, support, or enhance

health and healthy behaviours. Public will campaigns are used to "…legitimize or raise the importance of a social problem in the public eye as the motivation for policy action or change." (Coffman, 2002). Evaluation challenges, strategies, and measures are somewhat different for each type of campaign. Ideally, governmentled individual behaviour change campaigns will raise awareness, produce behavioural change, revise the social context within which behaviour occurs, and produce new demands on the government to further advance environmental shifts to reinforce and produce new behaviour change. For example, a government-sponsored campaign on the health risks of tobacco use could lead to public demands for government services to treat tobacco dependence, and a new tax on tobacco products to pay for the services. The public will campaign for a higher tobacco tax, and dedicated use of the new resources may be coordinated by nongovernmental organisations, but may eventually lead to a government-sponsored campaign to increase access to tobacco dependence treatment.

Individual behaviour change campaigns ("public education campaigns") emphasize advertising and marketing as a main strategy. Campaign planners and evaluators must have a clear sense of what the campaign will cause to happen, why it will happen, and who it will happen to, based on some theory of behaviour change (described below in the section). Measures of campaign effectiveness will center on what members of the target group will be aware of, know, and do as a result of the communication campaign that is different from what they were aware of, knew, and did before the campaign (National Cancer Institute, 2002), or that is different from what a comparable group is aware of, knows, and does related to topics addressed in the campaign.

Public will or engagement campaigns are used to build public demand ("will") to address a particular problem through policy and social action. Public will campaigns focus on the public's responsibility to create the supportive environment that will allow or promote a desired behaviour change (Coffman 2002; National Cancer Institute, 2005). The key strategies of public will campaigns are media advocacy and public relation, with reinforcing and supporting community action, including community organising and policy advocacy. Public will campaigns seek to set the public agenda by influencing the media agenda (and the way people and decision makers are exposed to and process issue information) through media advocacy. But the ultimate objective of policy or social change is achieved because the public will campaign prompts people to act, not by adopting a particular health behaviour, but by supporting (demanding) a particular policy change.

Public communication campaigns include a variety of communication, and other strategies, to educate the target population and disseminate information in compelling and engaging ways to raise the level of discomfort individuals have with a particular behaviour (e.g. tobacco use). They also pressure decision makers on specific issues for the purpose of changing (or advancing) policies. Types of public communication include paid (or "mass") media, public relations, media advocacy, and community action implemented discreetly or in combination (Coffman, 2002; Dorfman *et al.*, 2002). Thus, the public communication campaign components shown in Figure 5.33 can be implemented and evaluated as a multi-component intervention, with the interventions and outcomes in each "row" influencing outcomes in other rows, or as discreet campaigns, with outcomes following linearly from the specific intervention.

Paid media:

Paid or mass media is often the most expensive component of a public communication campaign, and yet may be the one that reaches the greatest number of people. It can be effective in communicating a tightly controlled message, creating an image, brand, theme, or call to action for the overall campaign, and can change attitudes, beliefs, and knowledge in the target population. Paid media, also known as advertising, introduces an issue or

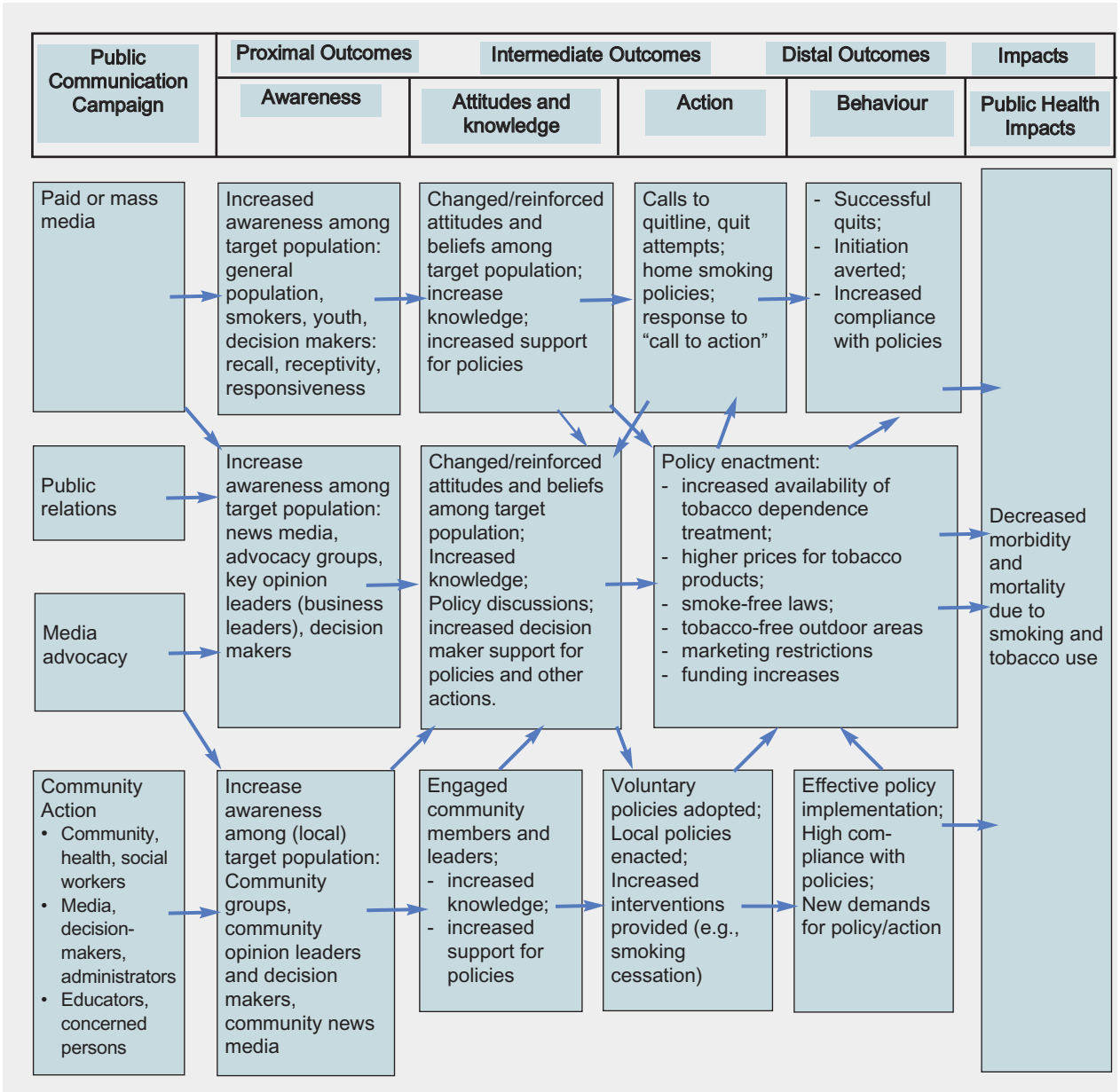


Figure 5.33 Flow diagram of public communication components and proximal and distal outcomes

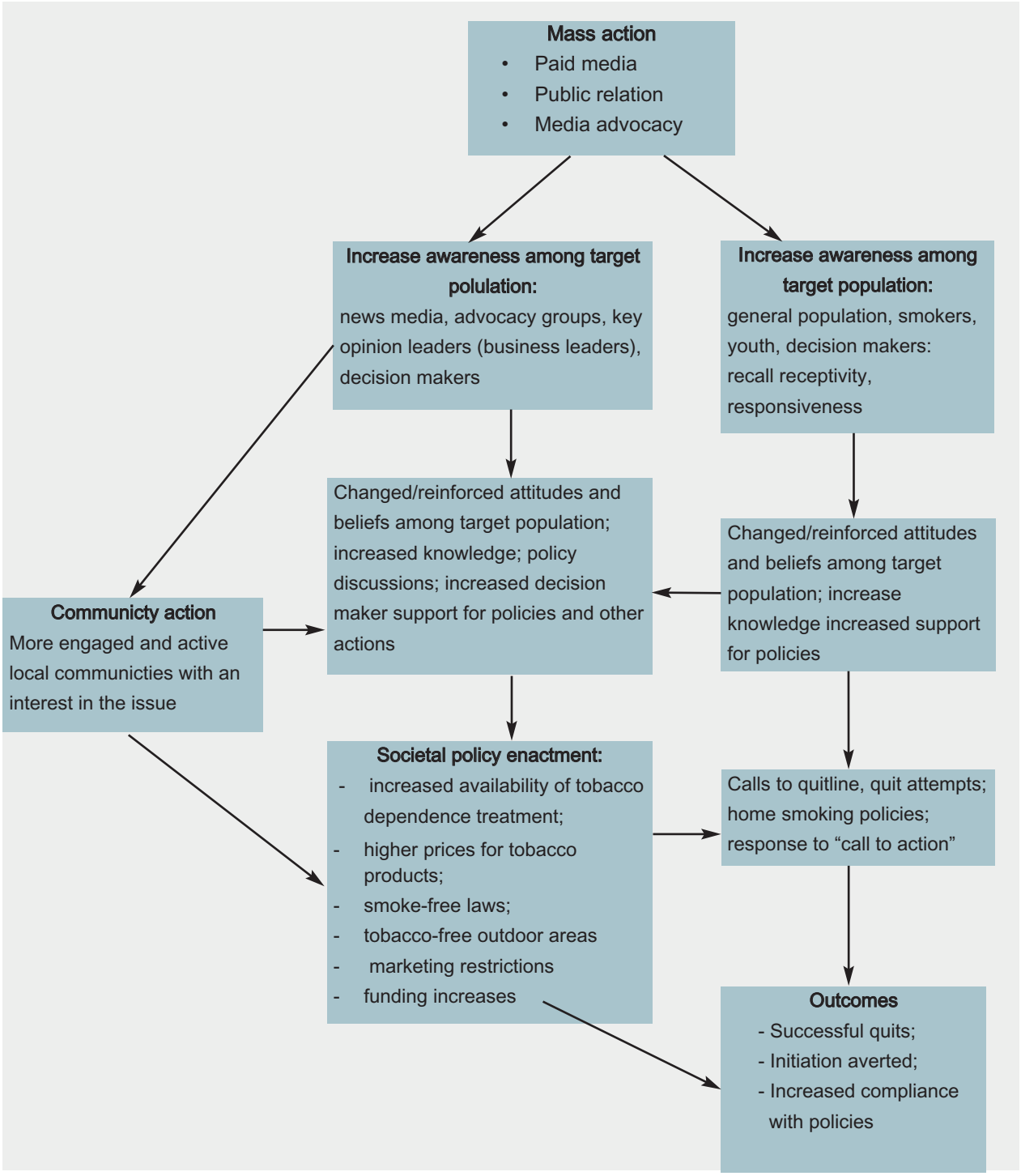


Figure 5.34 Conceptual framework for evaluation of anti-tobacco public communication campaigns

concept, delivers it to a large audience, and, if done effectively, raises awareness, increases knowledge, creates interest, engagement, concern, and stimulates conversation and action (Centers for Disease Control and Prevention, 2003). Paid media is not a necessary part of a public communication campaign. However, if resources are available, it can dramatically expand the reach of a campaign and reinforce and support the public relations and community action components. Paid media also may be used as a media advocacy strategy, with key messages strategically placed in print, electronic, and other media, to reach influential decision makers and opinion leaders, including policy makers. Mass media that is not paid for (e.g. media campaigns that rely on donated time and the use of public service announcements), can serve the same function as paid media in an overall public communication campaign, but is unlikely to have the reach of paid media or the target specificity; in addition, placement most likely will be outside the control of the campaign. Depending on the resources available, and the specific targets of the campaign message, paid media campaigns can feature a variety of media channels including television, radio, print, transit, billboards, Internet, brochures, and others.

Public relations:

The goal of public relations is to disseminate public communication campaign messages through others, specifically the news media, opinion leaders, and those who may be perceived as having more credibility or objectivity than campaign sponsors or paid media messages. Exposure obtained from public relations is "earned" coverage; "earned" because it is not paid for but obtained through strategic advocacy efforts, including working with news media outlets, community leaders, policy makers, and others with influence to disseminate key messages. Public relations provides opportunities to reach the target audience through sources that appear more legitimate, and allows the provision of more detailed information than paid media, all while positioning the campaign positively and potentially influencing the policy debate (Centers for Disease Control and Prevention, 2003). Public relations also provides the opportunity to "localize" national and international news, events, and research (Chapman & Dominello, 2001; Niederdeppe *et al.*, 2007), and bring to life local stories of personal tragedy (e.g. related to tobacco use) that can stand on their own or be coordinated with and reinforce paid media messages.

Public relations involves establishing relationships with members of the press, and other influential members of the community, developing supporting materials including press releases and press kits, and staging community events and press conferences, among other strategies. News and other media play a large role in determining what the public thinks about (agenda setting), how information is organised and packaged for public consumption (framing), and focuses the public on particular information at particular times for use in decision making (priming). Thus, public relations strategies are key elements of public communication campaigns and should be vigorously implemented as part of public engagement campaigns, in particular, that seek policy or social change (Wallack *et al.,* 1993, 1999; Coffman, 2002). While public relations strategies are employed to set the public agenda and keep issues in the public eye, they are often directed at specific policy makers and become part of a media advocacy strategy.

Media advocacy:

Media advocacy is an effort to use the tools of mass media and public relations to reframe the public debate, encourage a community to rethink its norms, and reach decision makers who have the power to transform the community environment through the adoption of policies that enhance public health (Wallack *et al.*, 1993, 1999; WHO, 2004). Media advocacy differs from paid media in that its main target is comparatively small (and could be only one individual), and the goal is policy change that will promote, support, or reinforce individual behaviour change and the public health agenda. However, media advocacy can use paid media as one strategy to accomplish advocacy objectives. In order to reach those individuals

with the power to make the policy change, media advocacy efforts can target highly organised and motivated individuals (or organisations) who can pressure policy makers to make the desired change. Media advocacy may even target the general public in an effort to set the public agenda and reframe an issue. In this case, paid media is a tool of media advocacy; communicating a message to policy makers through engaged citizens as the target audience. Mass media campaigns showcasing responsible tobacco company behaviour are likely media advocacy campaigns targeted at politicians and voters in an effort to recast the company's public image, earn the respect of the public, and relieve public pressure on policy makers to take action that would constrain the industry. Just as paid media, targeted at specific groups of individuals, may use a variety of messages that cajole, engage, cause fear, or provoke anger in an effort to stimulate behaviour change, media advocacy uses both positive and negative tactics to exert pressure on decision makers and provoke political action.

Community action:

In the context of public communication campaigns, community action engages the community in defining a problem locally and taking community-specific steps to advance a behavioural, normative, or policy shift at the local level or in support of state or national goals. Community action is linked to, and increases the resources of, the larger public communication campaign, raising awareness, engaging local news media, organising community events, disseminating information through local channels, and meeting with (and advocating with) local officials (Pierce *et al.*, 1990; Bracht, 2001). These community efforts are often legitimized, reinforced, and supported by paid media. Where paid media may not be possible, community action becomes a crucial component of public communication campaigns, often incorporating community organising tactics to advance media and policy advocacy objectives. Community action is both an extension of the public communication campaign to the local level, and a strategy in support of key public communication campaign components. It can take the form of community advocacy, public relations, participation in government processes, decision maker education, leadership training, staged events (e.g. press events, media advocacy, and grassroots mobilization), and community organisation to demand change (Niederdeppe *et al.*, 2007). Community action also increases the likelihood that the public communication campaign messages and results will endure long past the formal end of the campaign (Bracht, 2001).

Theoretical underpinnings

Grounding a campaign in one or more theories of behaviour change enables campaign planners to explain why and how a campaign should work, thus assessing the campaign's progress throughout the health communication process, not just at the end of the campaign (Atkin, 2001; Coffman, 2002; National Cancer Institute, 2002; Randolph & Viswanath, 2004). Assessing progress enables planners to improve the campaign as it is developed and implemented, before more resources have been invested in a campaign that may not succeed. Public communication campaigns that are grounded in theory are easier to evaluate over the lifetime of the campaign (and easier to causally link to outcomes), as planners are able to identify at the outset the more immediate or proximal indicators of whether a campaign is on track, as well as the longer-term indicators of campaign effectiveness. Change theories relevant to public communication campaigns include: the theory of reasoned action, social cognitive theory, the health belief model, the trans-theoretical model ("stages of change"), consumer information processing model, organisational change theory, community organisation theory, and diffusion of innovation theory (among others), each described briefly below. Readers are referred to Connell & Kubisch (1998), Atkin (2001), Bracht (2001), Coffman (2002), and the National Cancer Institute (2002) for additional information, bibliographies, and primary sources.

The theory of reasoned action postulates that attitudes and

norms create behavioural intentions, which in turn cause behavioural outcomes. A public communication campaign may be designed to change or reinforce specific attitudes and norms for the purpose of causing behaviour change. An evaluation of such a campaign would assess reinforcement of or shifts in attitudes and norms, and would only expect behavioural change where attitudes were or became consistent with the desired behaviour change. If such attitudinal shifts failed to occur or were not reinforced, the campaign would likely be revised.

Social cognitive theory postulates that behaviour change results from motivation to change and the acquisition of skills and abilities (self-efficacy) to change, within a given environmental context. A public communication campaign grounded in this theory would try to attract the target audience's attention, convey a compelling message, impart specific skills, and provide motivation to undertake behaviour change (preferably in conjunction with a reinforcing environmental change, such as a price increase on cigarettes, or the adoption of a smoke-free policy). An evaluation of such a campaign would assess attitudes and knowledge (skills) in the target population and desire to change the behaviour. In addition, a firm understanding of the environmental context would help shape the development of the campaign.

The health belief model suggests that people change behaviour when they feel susceptible or vulnerable as a result of a given behaviour, and believe that the costs of continuing the behaviour outweigh the costs of changing the behaviour.

The trans-theoretical model ("Stages of Change") posits that people proceed through (linearly or cyclically) a readiness continuum of behaviour change stages from pre-contemplation to maintenance of the behaviour change. Public communication campaigns based on this theory will identify the specific stages of the target population and attempt to move them to the next stage, will have different messages for audiences in the different stages, or, perhaps, will target people at one stage only. Evaluation outcome measures will be determined by the purpose and target audience of the campaign, and may be limited to shifts along the readiness to change continuum (e.g. from "happy to smoke" to "thinking about quitting").

The consumer information processing model suggests that how much and what kind of information people have and how they process it, are determinants of whether people will use information to inform and motivate behaviours or behaviour change. To increase the chances that information will be used in decision-making, public communication campaigns must make information available, package it as innovative and useful, and ensure that it is accessible to (able to be processed by) the target population. Tenets of this theory are particularly helpful for evaluating campaign messages, materials, and delivery media during the planning phases and early implementation to ensure that messages are understood by and resonate with the target audience.

The principles of community organisation theory are based on community empowerment and capacity building. In order to be successful and have a sustained impact, public communication campaigns must include partnerships with community members, organisations, and governments, and mobilize communities to develop and implement strategies in support of campaign goals. Evaluation of a campaign based on this theory would include stakeholder interviews, measures of community competence, monitoring of community activities, and other community evaluation techniques.

Diffusion of innovation theory describes how new norms, ideas, products, and practices diffuse through communities and become accepted or established in society. The theory focuses on characteristics of the innovation, as well as characteristics of the community, social networks, and communication systems through which the innovation is spread. Cigarette use is a primary example of how a new product "catches on" and diffuses through communities. Currently, smoke-free norms are being re-established, with the "diffusion" explained by this and other theoretical models.

Often, public communication campaigns are grounded in several theories in order to account for the complexities involved in behavioural and social change enterprises. They may even adopt new theoretical approaches as the campaign proceeds, based on evaluation findings, which might revise their understanding of the local (or audience-specific) change process, or provide new information about population attitudes and beliefs.

Theory and practice of public communication campaign evaluation and approaches to evaluating each component

In the context of the WHO FCTC, evaluation of public communication campaigns should assess whether the campaign is meeting its objectives as it is being planned, developed, and implemented in order to best ensure success, and demonstrate that the campaign has indeed achieved the expected outcomes. Thus, evaluation resources should be invested at the planning and developmental stages to ensure that specific interventions are customized to the target population and are culturally specific and appropriate. They should also be invested over the life of the campaign, and beyond, to assess whether proximal and more intermediate outcomes are being met. Particularly where innovative or unproven strategies are being implemented, or new theoretical models are being tested, more formal outcome studies may be appropriate. Public communication campaigns may be discreet interventions with a beginning, an end, and a predictable sequence of events in the middle. Often, however, they are more accurately described as a "messy social process" (Hornik, 2002), diffused by multiple strategies, through multiple channels, across individuals, communities, and institutions, with direct and indirect effects and diffuse outcomes that may reverberate long past the official "end" of the campaign (especially if policy change objectives were achieved) (Freimuth *et al.*, 2001). As a result, evaluation resources are appropriately invested in ongoing surveillance, point in time monitoring, special studies to identify opportunities for improvement, confirm that progress is being made, identify mediational and moderator effects (see Section 3.2), and link interventions to specific milestones and outcomes. Experimental designs and controlled trials often are not possible or appropriate (Balch & Sutton, 1997; WHO, 1998b), but instead a collection of information, existing data, and specific studies are needed to fully understand whether and how a campaign worked. Tightly linking campaign objectives to proximal outcomes can help demonstrate impact and, in particular, can help rule out competing explanations for observed change. Table 5.40 lists methods to assess the effectiveness of each public communication campaign component at various levels of evaluation, including establishing proximal outcomes.

At the outset of a public communication campaign, the problems and issues to be tackled and the baseline situation will have been established through ongoing surveillance or, at the community level, a needs assessment. The programmatic evaluation typically is conceptualized and implemented in four stages (described below) throughout the life span of the intervention and beyond. Formative evaluation begins as campaign concepts are being developed and summative evaluation focuses on the overall value of the campaign in terms of accomplishing its stated objectives. At the front end, evaluation includes testing and verification of campaign concepts ("formative") and careful monitoring of campaign activities and resources ("process") to ensure the campaign is being developed and implemented appropriately, efficiently, and with some likelihood of success (Atkin & Freimuth, 2001). At the back end, "outcome" evaluation answers the questions of whether the campaign has achieved its short- and long-term objectives, and has value to the community in terms of advancing public health goals (a major focus of this volume). The point is that evaluation should be well integrated into all phases of the public communication campaign, and this information should be well-utilized throughout the life

Communication Strategy ***Level of Evaluation***	**Paid media Public service announcements**	**Public relations and earned media**	**Media advocacy and government relations**	**Community action**
Ongoing Surveillance	*Structured analysis of data from existing surveillance systems*			
Formative				
Do the messages, materials, strategies "work?" Are they tailored to the intended audience?	Focus group discussions; Internet panel studies; Marketing surveys; Document analysis	Key informant in-depth interviews; document analysis	"Who do you know?" inventory; key informant in-depth interviews	Community needs assessment; community capacity analysis; strengths, weaknesses, opportunities, threats (SWOT) analysis; health risk profile
Process				
Implementation process: what and how much was done? Distribution, effort expended, resources committed	Gross rating points (GRPs)/Target rating points (TRPs), which are available from media buying firms and media channels, provide indicators of reach, frequency, exposure, and impressions.	News media tracking: count of stories run; tobacco control advocacy groups cited; content analysis, slant. Case study	News media framing analysis	Activity logs; meeting minutes
Proximal and Intermediate Outcomes				
Knowledge, attitude, policy, normative shifts	Population based/random digit dial (RDD) surveys (in-person, mail, telephone, Internet) of knowledge, attitudes, beliefs, behaviours; calls to telephone help line; web site visits, measures of responses to specific calls to action.	Special population surveys/key informant interviews; Official records of government policy and NGO policy.	Document analysis of legislative records. Case study.	Community policy database: voluntary, statutory/regulatory; Case study.
Distal Outcomes				
Behaviour change, disease rate change	Cigarette tax and sales records; behavioural risk factor and disease surveillance; disease registries; vital records			

Table 5.40 Methods to Assess the Effectiveness of Public Communication Campaigns by Campaign Component and Level of Evaluation

cycle of the campaign, and beyond, to make judgments about campaign progress, improve its effectiveness, and inform decisions about its future (Patton, 1997). Tables 5.41 and 5.42 list indicators of the effectiveness of each public communication component and corresponding outcomes by evaluation level. Each level is described more fully below.

Formative evaluation:

Formative research and evaluation identify the causal pathway through which an intervention is likely to work, and facilitates campaign improvement as it is being developed and implemented. It does not speak to the campaign's value or impact, but identifies its strengths and weaknesses, and aspects of the campaign that are not working as planned or are not likely to succeed (Mark *et al.*, 2000). It can provide information about key messages that are or are not resonating, and the types of individuals who are or are not responding to the campaign, among other variables important to its success. Information from formative evaluation is used by campaign planners and staff to solve problems, address weaknesses, revise expectations, revamp the campaign concepts and executions, or otherwise improve conceptualization and implementation (Patton, 1997). Evaluation and research, such as marketing surveys, that inform the creative process, also serve as baseline measures of attitudes, beliefs, and norms the public communication campaign is attempting to change.

Process evaluation:

Process evaluation is applied to programme implementation and answers the question how well the campaign is being delivered to the intended audience. Process tools measure effort and activity and help inform whether a campaign is being delivered as intended, and, if not, where the shortfalls are occurring. Retrospectively, process evaluation can shed light on what went wrong, if a particular campaign fails to meet its objectives, and identify lessons on how to make future campaigns more effective. Process evaluation involves monitoring resources, activities, and inputs including materials produced and distributed, news contacts made, meetings held, and a variety of information related to the placement of paid media. Process evaluation does not address the achievement of campaign outcomes or impacts, but can be used to link campaign activities to those outcomes by quantifying the "dose" of the campaign over time and in different communities.

Outcome evaluation:

Evaluation strategies for proximal outcomes are used by public communication campaign planners and evaluators to determine whether the shorter-term outcomes the campaign was designed to achieve have actually been met. As outlined in Figure 5.33, outcome evaluation generally requires more resources than formative or process evaluation, and, depending on the availability of financial and scientific resources, may be accomplished by special studies or by accessing information from routinely collected data sources. Outcomes of public communication campaigns vary from cognitive shifts (proximal) through social normative and behavioural shifts (distal), including individual knowledge, beliefs, awareness, attitudes, self-efficacy, behavioural intentions, behaviour through environmental changes, media frames, policy enactment, and normative change (measured policy enactment and compliance with policies).

Evaluation of more distal outcomes assesses achievement of public health goals, which almost certainly do not result from public communication campaigns alone. Impacts would include changes in health behaviours (e.g. tobacco use), tobacco-related disease rates (e.g. lung cancer incidence), and, ultimately, rates of death due to tobacco use. Outcome evaluation can be the most rigorous, complex, and resource intensive level of evaluation, and should be considered carefully at the programme (not the campaign) level. Public communication campaigns, after all, constitute only one component of the WHO FCTC effort to transform society and "reaffirm the right of all people to the highest standard of health."

Communication Strategy *Level of Evaluation*	Paid or mass media Public service announcements	Public relations and earned media	Media advocacy and government relations	Community action
Formative Do the messages, materials, strategies "work?" Are they tailored to the intended audience?	How is the message likely to make the audience feel? What message is understood and will the audience take away? What part of the message pleases, annoys, angers, scares the audience? Is the ad believable? Does the message speak to "people like me?" Is the message culturally appropriate? Is the message compelling? What is the appropriate channel for the message? What are the competing messages?	What is the current "information environment?" Will the message change the "information environment?" What *kind* of news to make (how to frame the message)?	Is the policy/ legislative environment hostile or hospitable to the message? What are the competing priorities? Who are allies? What are obstacles?	Strengths, weaknesses, opportunities, threats (SWOT) analysis results; meetings with community members and leaders; formation of community advisory group
Process Implementation process: what and how much was done? Distribution, effort expended, resources committed	Number of ads running, placement, impressions, Gross rating points (GRPs)/ Target rating points (TRPs), money spent, location of out-of-home media, time lines met	News media tracking: count of stories run; tobacco control advocacy groups cited; case study; content analysis; framing analysis (point of view, accuracy, slant, agenda setting)	Indicators of decision maker interest and action from public hearings and official meetings	Community meetings held; coalitions formed; organisations involved; number of activities, trainings and events planned/implemented;number of people who participate; resources invested in outreach (money, time, personnel); number of materials produced and distributed

Table 5.41 Formative and Process Indicators of the Effectiveness of Public Communication Campaigns by Campaign Component and Level of Evaluation

Communication Strategy *Level of Evaluation*	Paid media Public service announcements	Public relations and earned media	Media advocacy and government relations	Community action
Short-term Outcomes Awareness, knowledge, Attitude shifts	Confirmed awareness (discrimination) of media message; level of receptivity to media message (e.g. talked to others about it); support for specific policy: increased availability of tobacco dependence treatment; higher prices for tobacco products; smoke-free air laws; marketing restrictions; funding increases	Did the issue get on the media agenda? Was the issue framed according to the campaign objectives? Did the media coverage advance the message? Public/decision maker support for specific policy	Did the issue get on the public agenda? Support for policies; legislative proposals submitted; legislative bills introduced	Voluntary policies adopted; health care policies to provide tobacco dependence treatment; better informed professionals; improved health care services; availability of cessation support
Intermediate Outcomes Knowledge, attitude/policy shifts	Number of responses to call to action: calls to quitline, visits to web site, other); knowledge and attitude shifts: reduced acceptability of smoking/ exposure to tobacco smoke; increased awareness of harm from smoking/ tobacco smoke; increased intentions to quit; increased knowledge of how to quit	Policy enactment	Policy enactment; amount of cigarette or other taxes.	Community laws and regulations enacted; community services and programmes established
Distal Outcomes				
Behaviour normative change, disease rate change	Per capital consumption of cigarettes; smoking prevalence, use of other tobacco products; exposure to tobacco smoke; incidence of tobacco caused disease			

Table 5.42 Outcome Indicators of the Effectiveness of Public Communication Campaigns by Campaign Component and Level of Evaluation

If done well, they will most likely contribute to population or target group changes in, or reinforcement of, attitudes, knowledge, and beliefs that contribute to policy, environmental, and normative improvements that advance the public health.

Measures to assess proximal and distal outcomes

Of all the public communication campaign components, evaluation indicators are probably the most highly developed, or as least the most familiar, for paid media. Indicators include results of focus group testing of media messages to ensure they "speak" to or "resonate" with the target audience, convey the intended message, and are likely to provoke the desired attitude and behaviour changes. Indicators of the campaign's reach into the target population and the frequency with which campaign messages were aired (usually quantified as gross rating points (GRP) or target rating points (TRP) for television ads and viewer "impressions" for print media) are common process measures. Reach also can be quantified by means of consumer surveys designed to elicit awareness of the campaign (i.e. aided or unaided recall of specific campaign ads, messages, and themes). Proximal measures of campaign effects, for example changes in awareness, attitudes, and knowledge about the issue being promoted, can be obtained by in-person, mail, telephone, or Internet surveys of the target population, where outcome measures can be linked to awareness. More distal outcomes, like actual behaviour change, also are typically measured by some kind of survey of the target population. Similar survey tools and methods are used to assess population support for specific policy initiatives. Depending on the purpose of the campaign, other measures of campaign effectiveness may be appropriate. For example, indicators of effectiveness of a paid media campaign designed to promote telephone-based cessation services could include the number of calls to a helpline or number of calls among those aware of the campaign. Paid media campaigns promoting other calls to action (e.g. to visit a web site, sign a petition, or send a letter), would be similarly evaluated by the number of individuals who respond by taking the requested action. Systems would have to be established to compile and count these actions, and may be as simple as monitoring "hits" to a web site before and after a call to action, tallying the number of signatures on a petition, or including postage paid (addressed) envelopes in a letter writing campaign, with the postage charge providing information on the number of letters sent.

For process and outcome measures, in order to support a claim that changes in awareness, attitudes, and behaviours result from the campaign itself, evaluators need to demonstrate that the effects occurred uniquely, temporally, or to a greater extent in the target population. This may be achieved by identifying a comparison community not receiving the public communication campaign. This could either be the target population, prior to implementation of the paid media campaign (good), a similar community not receiving the campaign at the same time as the target population (better), or by varying the dose of the campaign across jurisdictions (Farrelly *et al.*, 2005b). In some cases, intercept surveys, or surveys of available members of the target population ("convenience" samples), will be the only practical means of gleaning the potential impact of a campaign. Such surveys may be useful for obtaining anecdotal information, identifying problems, or fleshing out the details of findings from larger population surveys, but generally are not considered to be robust indicators of population level outcomes.

Perhaps the most common indicators of effective public relations result from news media tracking. This is both a simple count of news stories related to the tobacco topic promoted by a public relations effort (or a ratio of such stories to other health-related stories), and a content analysis of those articles to determine characteristics such as the message, accuracy, slant, point of view, and prominence of message, among others (Henry & Gordon, 2001; Durrant *et al.,* 2003; Clegg-Smith *et al.*, 2005; Neiderdeppe *et al.*, 2007). It may

even include responses to the story, such as letters to the editor, news media-sponsored Internet polls, and whether the story was "picked up" by other media channels and outlets. Together these indicate intermediate outcomes; demonstrating first that the public relations efforts resulted in increased news media coverage, and second that the coverage conveyed the information and point of view promoted by the public relations campaign. News media tracking efforts typically are limited to print news media, which has been shown to be a marker of media coverage overall (e.g. electronic media) (National Cancer Institute, 2006). However, coverage of television media can be tracked as well, with volunteers or paid viewers systematically viewing and cataloging television news coverage of specific issues for content and characteristics. While counting news stories and describing characteristics relevant to tobacco control is straight-forward, connecting them to specific programmatic outcomes may be more difficult, as noted by evaluators of the American Stop Smoking Intervention Study:

> *The challenge in evaluation, however, is demonstrating that news media coverage does in fact influence the thinking, decisions, and behaviour of the public and of policy makers. Although determining such a cause-and-effect relationship for some very focused and geographically limited topics might be possible, researchers in the field of evaluation are still grappling with how to do so for wide-scale public health interventions* (National Cancer Institute, 2006).

Process indicators become particularly important in making the link, as well as understanding how some strategies might be improved in the event that anticipated results are not achieved. Anecdotes and personal statements may be particularly relevant to understanding the influence of news stories on decision maker action.

The longer-term goals of the WHO FCTC are to change individual behaviour as a result of modifications to the social, economic, and health environment, which in turn result from government intervention (WHO, 2003). These modifications provide the conditions within which people can be healthy (Institute of Medicine, 1988). The implementation and success of these interventions are based in part on popular expectations and demands. Media advocacy strategies put these policy change debates on the public and policy maker agendas. By focusing media attention on specific public health issues (agenda setting), and focusing the debate to reflect the public health perspective (framing), media advocates seek to influence the information the public uses to make decisions (priming), and reach opinion leaders and policy makers to change public policy (Wallack *et al.,* 1993). Indicators of media advocacy success include: measures of whether the campaign issue has become part of the media agenda and is framed according to the public health perspective (established through content analysis of news media programming and print news articles), whether media support the particular policy agenda (e.g. in newspaper editorials) and whether their support is associated with policy maker support (e.g. through key informant interviews), and whether that particular agenda is advanced (e.g. in legislative debate, the introduction of legislative bills or the enactment of legislation). Areas of exploration for formative and process evaluation of media advocacy include: an assessment of the message's connection with people at the community level, the media's understanding of the issue, how the issue can be framed to capture media and public attention and focus attention on larger public health values (e.g. how the tobacco control issue can be framed to emphasize social accountability rather than personal responsibility (Wallack & Dorfman, 2001)), how relationships have been developed with community advocacy groups and the media, and a quantification of the actions of these groups. Measures of proximal outcomes associated with media advocacy include: public support for specific policy goals as measured by in-person, mail, telephone or Internet

surveys, and political polling, and framing analysis to monitor and assess news media reports of specific policy initiatives and of the issue or problem the policy initiative seeks to address. Standard methods include: key informant and opinion leader interviews, political polling, news media tracking, and content analysis. Since the key targets of media advocacy are the policy makers (i.e. organisations or legislative bodies that have the power to make the policy change), indicators of success will be drawn from official meeting minutes and transcriptions, key opinion leader interviews focusing on specific initiatives, and official records of policy enactment, as well as news media reports.

Public communication campaigns are most effective in changing community and social norms, and building support for and actually prompting the enactment of public health policy, when they incorporate community action (Hopkins *et al.*, 2001). However, what constitutes community action, how community initiatives are described, quantified, and measured, and what change theories underlie their development and success, have been topics of ongoing debate (Connell & Kubisch, 1998; Bracht, 2001; Connell *et al.,* 2007). Successful community action involves change at many levels (individual, family, personal network, institutional, and community), including many domains (economic, social, physical, and community), and evolves over different time periods (near-term, interim, long-term, or ultimate) (Connell & Kubisch, 1998). Thus, the measurement challenge is substantial. As with public communication campaigns overall, evaluation of community initiatives is easier when an explicit change theory is specified at the outset from which to identify specific indicators of the development and implementation of the initiative and progress toward anticipated short-, intermediate, and longer-term outcomes. Indicators should reflect the processes through which activities are developed and planned (community needs assessment, meetings held, individuals present, organisations involved), the implementation of those activities (e.g. meetings with decision makers, community forums, press events), and proximal results of these activities (the adoption of resolutions, community participation in events, news coverage, improvements in awareness of problems and solutions, increase in community member knowledge about the specific issue and support for specific action, evidence of decision-maker support), as well as longer-term change (policy enactment or the achievement of a specific objective, such as provision of community cessation services or removal of pro-tobacco advertising at a specific location) (Gambone, 1998). Methods for describing, monitoring, and capturing the effects of community action, community interventions, and measures of short- and longer-term outcomes are only poorly developed at this time. Nonetheless, cataloging the input that supports community action (financial, in-kind, and personnel resources), quantifying activity levels (number of meetings convened, contacts made, events held), and documenting process through case studies, can be helpful in discerning whether and how community actions contribute to public communication goals.

Table 5.40 summarizes methods to assess the effectiveness of public communication campaigns by campaign component (paid media, public relations, media advocacy, and linked community action), and level of evaluation (formative, process, outcome), as described in this section. Not all jurisdictions will have the resources to implement a multi-component public communication campaign or field population-based surveys to assess campaign outcomes. For such jurisdictions, the methods within each column may be used discreetly for each level of evaluation within a particular component. It is not necessary to mount a population level survey in order to demonstrate a population level impact, but some population level data base (like emergency room admissions for acute myocardial infarction, calls to a quitline, or sales of cessation medication), with information proximally related to the campaign result, is needed.

Tables 5.41 and 5.42 list formative and process indicators and proximal and distal outcome

indicators, respectively, of the effectiveness of public communication campaigns, by campaign component and level of evaluation. Evaluation indicators of proximal campaign effects should be specific to the individual campaign message and communication component. For example, survey questions designed to understand the target audience's awareness of, or reactions to, a set of commercials or advertisements describing the health risks of exposure to secondhand smoke, will be specific to the content of the message and the goals of the campaign (e.g. to increase support for the enactment of or compliance with a smoke-free workplace law), and will be different from survey questions associated with a campaign to promote cessation among current smokers. In both cases, campaign planners and evaluators should test the messages with members of the target audience, monitor implementation of the campaign, assess exposure to and understanding of the messages, determine attitudes, beliefs, and knowledge related to the topic, before and after the campaign (or among those exposed and not exposed to the campaign), and assess changes (increases) in the likelihood of engaging in the particular campaign "call to action" (e.g. refrain from smoking in public places, demand no smoking in public places, think about quitting smoking, make a quit attempt). The specific questions used will be determined by the content and goals of the public communication campaign. Examples of specific questions are provided below.

A helpful source of indicators to measure outcomes associated with public communication campaigns is the Centers for Disease Control and Prevention (CDC) manual, *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programmes* (Starr *et al.*, 2005). This user-friendly, fairly comprehensive guide compiles and provides information on 120 outcome indicators for use in evaluating the short-, intermediate, and longer-term impacts of comprehensive tobacco use prevention and control programmes. Indicators are organised according to three programmatic areas (preventing initiation, promoting cessation, and eliminating exposure to tobacco smoke), and grounded in evidence-based logic models. Detailed information is provided for each indicator including indicator definition, example data sources, specific measures (e.g. question wording), and overall quality of the indicator. Those indicators useful for monitoring the outcomes of public communication campaigns, specifically, are listed in Table 5.43. The guide is available online (in English) at http://www.cdc.gov/tobacco/Indicators/KeyIndicators.htm.

The Question Inventory on Tobacco, formerly known as The Survey Questionnaire Design Resource (available online (in English) at http://apps.nccd.cdc.gov/QIT/), is another important resource for identifying survey questions that contribute to effective measures of intervention outcomes. Also developed by the CDC's Office on Smoking and Health, the online resource categorizes more than 1700 tobacco-related questions from 13 United States national and state surveys that have been used starting in 1990. The Question Inventory on Tobacco resource provides easy-to-use search capabilities to locate survey questions, including possible answer formats, and identifies the specific surveys in which the questions have been used.

The WHO FCTC Article 12 describes five topics about which the public should be made aware through public communication tools: 1) health risks of tobacco consumption (including addiction), 2) health risks of tobacco smoke, 3) benefits of quitting, 4) aspects of the tobacco industry, and 5) adverse health, economic, and environmental consequences of tobacco production (this topic is not addressed in this section). In addition, Article 12 specifically addresses awareness of tobacco issues among media professionals, decision makers, community health and social workers, educators, and concerned individuals. Various strategies can be used to achieve public awareness, from traditional paid media campaigns utilizing television, radio and/or print targeting the general population or population subgroups (like smokers or youth), to strategic, targeted public relations and community action (or education)

Outcome Level	Indicator Number	Indicator description
Short-term	1.6.1	Level of confirmed awareness of anti-tobacco media messages
	1.6.2	Level of receptivity to anti-tobacco media messages
	1.6.4	Level of support for policies
	1.6.5	Level of support for increasing excise tax on tobacco products
Intermediate	1.10.1	Proportion of young people who think that smoking is cool and helps them fit in
	1.10.2	Proportion of young people who think that young people who smoke have more friends
	1.10.5	Proportion of young people who are susceptible never-smokers
	1.12.1	Amount of tobacco product excise tax
	3.11.1	Proportion of adult smokers who have made a quit attempt
	3.11.2	Proportion of young smokers who have made a quit attempt
	3.11.3	Proportion of adult and young smokers who have made a quit attempt using proven cessation methods
Longer-term	2.7.1	Proportion of the population reporting exposure to secondhand smoke in the workplace
	2.7.2	Proportion of the population reporting exposure to secondhand smoke in public places
	2.7.3	Proportion of the population reporting exposure to secondhand smoke at home or in vehicles
	3.13.1	Proportion of smokers who have sustained abstinence from tobacco use
	3.14.1	Smoking prevalence
	3.14.2	Prevalence of tobacco use during pregnancy
	3.14.3	Prevalence of postpartum tobacco use
	3.14.4	Per capita consumption of tobacco products

From Starr *et al.* (2005; http://www.cdc.gov/tobacco/Indicators/KeyIndicators.htm)

Table 5.43 Example Indictors from *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programmes* Relevant to Monitoring the Effects of Public Communication Campaigns

campaigns targeting community leaders, health care providers, business leaders, nongovernmental organisation directors, tribal leaders, and others. Specific measures to ascertain population level awareness and knowledge of topics included in Article 12 are described below.

Awareness of paid media campaigns by the general population or specific targeted subgroups (e.g. smokers, youth) is generally ascertained by population level surveys, including telephone surveys (care should be taken with written surveys to ensure that responses are not cued by response categories or other prompts). Awareness can be ascertained by general questions

that require the participant to fill in details of the campaign, or of specific ads, or by providing the participant with some general information about a campaign or ad in an effort to prompt or facilitate recall (Table 5.44). Unaided recall is generally considered to be a superior method for accurately estimating exposure to and awareness of a campaign or message (Sly *et al.,* 2001b), but aided recall has been shown to be an effective measure as well (Niederdeppe, 2005). Measures of awareness are designed to determine exposure to a specific message or advertisement, so that respondents can be categorized according to exposure level, and differences in attitudes and behaviours can be correlated to awareness. Some more general measures of awareness are intended to ascertain the amount of anti-tobacco advertising to which subjects are exposed, in order to make more general inferences about the relationship between anti-tobacco messaging in general, and attitudes and behaviours in general. Finally, measures of the relevance or salience of the media message to the individual provides key information on whether the campaign is effectively communicating the message, the types of individuals who are more likely to respond to the message, and the utility of the campaign in contributing to programme goals. This information can be used to strengthen a poorly performing campaign in progress or, at least, provide useful information for developing the next campaign. In order to assess exposure to and awareness and salience of messages to survey participants, measures need to be customized to the specific media campaign.

Once exposure, awareness, and salience are assessed and quantified, an analysis can be undertaken as to whether those who were exposed to the campaign message were aware of it or receptive to it in some way, are more likely to be aware of and understand the key messages of the campaign, and whether this new awareness or knowledge is associated with specific proximal outcomes (e.g. attitudes and beliefs). Measures of awareness and knowledge of specific campaign messages should be constructed to closely match specific messages being delivered and the overall intent of the communication campaign. The measures described in Table 5.45 relate specifically to awareness and knowledge of the topics described in Article 12. For campaigns that address other topics (e.g. motivating tobacco users to quit, issues related to "light" and "low tar" cigarettes, increasing support for specific policies), readers are referred to surveys and measures described in this Handbook and elsewhere (National Cancer Institute, 2002; Centers for Disease Control and Prevention, 2003; Starr *et al.*, 2005). Measures of behaviours (e.g. quitting, uptake, abstinence) that a campaign may seek to influence, as well as mediators and moderators of these behaviours (e.g. perceptions of risk, cooccurring disorders), are discussed in Section 3.1.

Measures of effectiveness for public relations, media advocacy, and community action efforts are less well developed and generally have not been collected and catalogued in the form of surveys, interviews, and question lists. Readers are referred to Tables 5.41 and 5.42 for listings of the types of information that should be collected in order to assess the effectiveness of these efforts. Additional information, resources, and "how to" instructions for assessing the effects of public relations and media advocacy efforts, in particular, are available from Radke (1998), National Cancer Institute (2002), Centers for Disease Control and Prevention (2003) and the WHO (2004), among others.

Monitoring other communication campaigns

Public communication campaigns constitute an effort to control the information environment (Randolph & Viswanath, 2004), to make specific information available to the target population, influence the public agenda, and frame the policy debate from a public health perspective, with the objective of changing behaviours, norms, and policies to advance public health. However, public communication campaigns often take place in a cluttered and competitive environment. They are competing for attention with other communication campaigns,

Construct	(a) Awareness of Specific Anti-Tobacco Media Messages
Measure	"Have you recently seen an anti-smoking or anti-tobacco ad on TV that shows-- [brief description of ad]?" (Yes, Maybe, No) "What happens in this advertisement?" "What do you think the main message of this ad was?"
Source	LMTS, 2003 (http://americanlegacy.org)
Validity	Established validity (Sly *et al.*, 2001b)
Variations	Variations are possible in the amount of prompting provided to the respondent (e.g. "Are you aware of any advertising or campaign against smoking or about or against cigarette companies that is now taking place?" (Yes, No) from [source?]) "What is the theme/slogan of this advertising or campaign?"
Comments	It may be necessary to assess overall TV viewing/radio listening patterns to understand whether participants had the opportunity to be exposed to the media message. Expected response categories should be pre-determined, but should not be read to the respondent. Responses are categorized as accurately describing the ad (indicating awareness) or not.
Construct	**(b) Awareness of General Anti-Tobacco Media Messages**
Measure 1: Adult	"Now I would like you to think about advertising or information that talks about the dangers of smoking, or encourages quitting. In the last 6 months - [since…] - how often, if at all, have you noticed such advertising or information?" (Never, Rarely, Sometimes, Often, Very often)
Source	The ITC Project, 2007
Validity	Face validity.
Variations	The time period of interest, the medium specified, and the types of ads described can all vary (e.g. "During the past 7 days, how many commercials have you seen on TV about NOT smoking cigarettes?" (None, One, Two or three, Four to six, Seven or more), from Global Adult Tobacco Survey (GATS, 2007). In this question, one can substitute (or add) "heard on the radio" or "seen on a billboard," as appropriate, for "seen on TV."
Comments	The use of this general item is helpful to characterize level of exposure to the broad range of state/provincial and national or other media-based anti-tobacco education campaigns. Such questions may be particularly helpful for pre-campaign surveys to quantify the amount of "background" anti-tobacco advertising to which the population is exposed.

Table 5.44 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Paid or Mass Media Components

Measure 2: Youth	"During the past 30 days (one month), how many anti-smoking media messages (e.g., television, radio, billboards, posters, newspapers, magazines, movies) have you seen or heard?" (A lot, A few, None)
Source	GYTS, 2007
Validity	Face validity.
Variations	"When you go to sports events, fairs, concerts, community events, or social gatherings, how often do you see anti-smoking messages?" (I never go to sports events, fairs, concerts, community events, or social gatherings, A lot, Sometimes, Never)
Comments	The use of this general item is helpful to characterize level of exposure to the broad range of state/provincial and national, or other media-based, anti-tobacco education campaigns. Such questions may be particularly helpful for pre-campaign surveys to quantify the amount of "background" anti-tobacco advertising to which the population is exposed. The variation may be useful for assessing awareness of general anti-smoking messages in non-electronic venues.
Measure 3: Locations	"In the last 6 months, have you noticed advertising or information that talks about the dangers of smoking, or encourages quitting in any of the following places?" (Yes, No) – READ OUT EACH STATEMENT • on television • on radio • at the cinema [US/Canada/AUS: at the movies] • on posters or billboards • in newspapers or magazines • on shop/store windows or inside shops/stores where you buy tobacco • on cigarette packs • leaflets • on the Internet • anywhere else? (specify)
Source	The ITC Project, 2007
Validity	Face validity.
Variations	Locations listed should be relevant to the campaign and the jurisdiction and should vary, as appropriate.
Comments	Understanding where consumers are exposed to anti-tobacco media may help in planning a public communication campaign, or may help identify specific ads or campaigns to which consumers have been exposed.

Table 5.44 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Paid or Mass Media Components

Construct	(c) Awareness of Smoking Related News Stories
Measure	"Now I want to ask you about the media more generally. First, thinking about news stories relating to smoking or tobacco companies that might have been on TV, radio, or in the newspapers. In the last 6 months, that is, since [6 month anchor], about how often, if at all, have you seen or heard a news story about smoking?" (Never, Rarely, Sometimes, Often, Very often) "On balance, how did the news stories portray smoking? Were they All pro-smoking, Mostly pro-smoking, Equally pro- and anti-smoking, Mostly anti-smoking, All anti-smoking?"
Source	The ITC Project, 2007
Validity	Face validity.
Variations	
Comments	These questions can help assess the effects of public relations and media advocacy efforts, when compared over time and referencing periods of campaign activity. Results should be cross referenced with news media tracking to better understand how people's perceptions correspond to actual reporting.
Construct	**(d) Salience of the Anti-Tobacco Media Message**
Measure 1: Adult	"This ad said something important to me. Would you say you… (Strongly agree, Agree, Disagree, Strongly disagree, Have no opinion, Don't know?" "After seeing this ad, did you talk to anyone about not smoking?" (Yes, Maybe, No)
Source	Wakefield *et al.,* 2003b
Validity	Established validity.
Variations	This question set focuses on not smoking. Depending on the content and purpose, other topics could be inserted in place of "not smoking."
Comments	This type of question is used following the respondent's description of a specific ad to gauge whether the respondent found the ad to be salient to his or her situation, and whether the ad prompted the respondent to think more about the topic.
Measure 2: Youth and young adult	"On a scale from 1 to 5, where 1 means you don't like this ad at all and 5 means you like the ad very much, how much do you like this ad?" (One, Two, Three, Four, Five) "Would you say the ad grabbed your attention?" (Yes, No) "Did you talk to your friends about this ad?" (Yes, No)
Source	Legacy Media Tracking survey (LMTS; http://americanlegacy.org)

Table 5.44 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Paid or Mass Media Components

Validity	Established validity. Thrasher *et al*., 2006b.
Variations	"Tell me how much you agree or disagree with the following statement: This ad is convincing. Would you say you… (Strongly agree, Agree, Disagree, Strongly disagree, Have no opinion, Don't know?" "Would you say the ad gave you good reasons not to smoke?" (Yes, No) "Would you say the ad makes you question the motives of cigarette companies?" (Yes, No) "Did you talk to your friends about this ad?" (Yes, No) From the Legacy Media Tracking Survey
Comments	The selection of appropriate questions for measuring salience depends on what the study is most interested in understanding. Question sets, as opposed to individual questions, are typically necessary to understand this construct. The examples provided here measure slightly different issues: was the message noticed and did it "create a buzz" versus did the message impart information that was integrated into the respondents thinking on the topic.

LMTS: Legacy Media Tracking Survey
ITC: International Tobacco Control Policy Evaluation Study
GATS: Global Adults Tobacco Survey
GYTS: Global Youth Tobacco Survey

Table 5.44 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Paid or Mass Media Components

typically commercial advertising and marketing campaigns, and they are competing for salience, relevance, and resonance with other efforts to promote behaviours and norms and control the terms of the policy debate. The success or failure, and the relative impact, of a public communication campaign will be dependent to some degree on what is going on in the larger information environment. This is particularly important for public communication campaigns focused on tobacco control issues. Tobacco and pharmaceutical companies use the same communication strategies, including paid media, public relations, media advocacy, and (to some extent) community action to promote their products and perspectives. Monitoring and understanding the larger information environment allows public communication campaigns to adapt strategies to respond to or reflect the realities of this environment and better understand and document the challenges and constraints that threaten the success of a public communication effort.

Tobacco and pharmaceutical company communication campaigns can be monitored with many of the same tools and indicators as a public communication campaign. However, key steps, processes, and information will be unavailable to public communication campaign planners and evaluators, such as the exact target and objectives of the campaign. For example, the target may appear to be youth, but is actually voting adults; it may appear to be smokers, but is actually policy makers; it may appear to be concerned adults, but is actually potential members of the jury. The objectives may appear to be preventing youth initiation or promoting adult

Construct:	(a) Health Risks of Tobacco Consumption
Measure 1: Awareness Adults	"To what extent, if at all, has smoking damaged your health?" (Not at all, Just a little, A fair amount, A great deal) **"How worried are you, if at all, that smoking** will **damage** your **health in the future?**" "**To what extent, if at all**, has **smoking lowered** your **quality of life**? How worried are you, if at all, that smoking will lower your quality of life in the future?"
Source	The ITC Project, 2007
Validity	Face validity.
Variations	
Comments	
Measure 2: Awareness Youth	"Do you think cigarette smoking is harmful to your health?" (Definitely not, Probably not, Probably yes, Definitely yes)
Source	GYTS, 2007
Validity	Face validity.
Variations	
Comments	
Measure 3: Knowledge Adults	"I am going to read you a list of health effects and diseases that may or may not be caused by smoking cigarettes. Based on what you know or believe, does smoking cause the following (Yes, No to each question): • heart disease in smokers • stroke in smokers • impotence in male smokers • lung cancer in smokers • lung cancer in nonsmokers from secondhand smoke?"
Source	The ITC Project, 2007
Validity	Face validity.
Variations	As far as you know, are each of the following chemicals included in cigarette smoke? (Yes, No) • cyanide • mercury • arsenic • carbon monoxide?" From the International Tobacco Control Policy Evaluation Survey

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

Comments	Depending on the purpose of the campaign and the selected key messages, questions will need to be modified to be relevant. As a baseline indicator, prior to campaign implementation (or as general surveillance), these measures estimate population level knowledge of health risks (but not perceptions of personal risk).
Measure 4: Knowledge Youth	"It is safe to smoke for only a year or two, as long as you quit after that? Would you say you… Strongly agree, Agree, Disagree, Strongly Disagree, No opinion?"
Source	Legacy Media Tracking Survey (LMTS, (http://americanlegacy.org)
Validity	Face validity.
Variations	"Do you think it is safe to smoke for only a year or two as long as you quit after that?" (Definitely not, Probably not, Probably yes, Definitely yes) From the Global Youth Tobacco Survey
Comments	
Measure 5: Addiction Awareness Adults	"Do you consider yourself addicted to cigarettes?" (Not at all, Yes–somewhat addicted, Yes–very addicted)
Sources	The ITC Project, 2007
Validity	Face validity.
Variations	
Comments	
Measure 6: Addiction Awareness Youth	If you started smoking regularly, do you think you could stop smoking anytime you wanted?" (Definitely yes, Probably yes, Probably not, Definitely not, No opinion)
Source	Legacy Media Tracking Survey (LMTS, (http://americanlegacy.org)
Validity	Face validity.
Variations	"Once someone has started smoking, do you think it would be difficult to quit?" (Definitely not, Probably not, Probably yes, Definitely yes) Source: Global Youth Tobacco Survey
Comments	

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

Construct	(b) Health Risks of Tobacco Smoke Exposure
Measure 1: Awareness Adults	"Do you think that breathing smoke from other people's cigarettes is… Very harmful to one's health, Somewhat harmful to one's health, Not very harmful to one's health, Not harmful at all to one's health, Don't know/Not sure?"
Source	GATS, 2007
Validity	
Variations	
Comments	The perception that environmental tobacco smoke (ETS) is harmful can be an important factor for gauging public support for tobacco control efforts. This question also can be an indicator of the effects of ETS education efforts.
Measure 2: Awareness Youth	"Do you think the smoke from other people's cigarettes is harmful to you?" (Definitely not, Probably not, Probably yes, Definitely yes)
Source	GYTS, 2007
Validity	Face validity.
Variations	
Comments	The perception that environmental tobacco smoke (ETS) is harmful can be an important factor for gauging public support for tobacco control efforts. This question also can be an indicator of the effects of ETS education efforts.
Measure 3: Knowledge Adults	"Would you say that breathing smoke from other people's cigarettes causes… (Yes, No to each question) [RANDOMIZE ORDER] …Lung cancer in adults …Heart disease in adults …Colon cancer in adults …Respiratory problems in children …Sudden infant death syndrome?"
Sources	CDC Adult Tobacco Survey; 1987 National Health Interview Survey (http://www.cdc.gov/nchs/nhis.htm)

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

Validity	Face validity.
Variations	
Comments	These items can gauge the level of public understanding of the health effects of tobacco smoke on nonsmokers. Colon cancer is included in this series as an indicator for "over-reporting" in order to estimate the possible magnitude of over-reporting.
Measure 4: Support for Policy Adults	• In the indoor dining area of restaurants • In indoor shopping malls • In public buildings • In bars and cocktail lounges • In day care centers • In indoor sporting events and concerts "… do you think that smoking should be allowed in all areas, some areas, or not allowed at all?"
Source	GATS, 2007
Validity	Face validity.
Variations	
Comments	Programmatic focus and activities, the goals of the communication campaign, and the local jurisdiction will determine which environments need to be included in the survey and whether additional environments are added. Such questions provide information on attitudes towards restrictions on exposure to secondhand smoke; a measure of social norms.
Measure 5: Support for Policy Youth	"Are you in favour of banning smoking in public places (such as restaurants, buses, "streetcars, trains, in schools, on playgrounds, in gyms and sports arenas, discos)?" (Yes, No)
Source	GYTS, 2007
Validity	Face validity.
Variations	
Comments	
Construct:	**(c) Benefits of Quitting**
Measure 1: Adults	"How much do you think you would benefit from health and other gains if you were to quit smoking permanently in the next 6 months?" (Not at all, Slightly, Moderately, Very much, Extremely)
Source	The ITC Project, 2007
Validity	Face validity.

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

Variations	"If a person has smoked a pack of cigarettes a day for more than 20 years, there is little health benefit to quitting smoking." (Strongly agree, Agree, Disagree, Strongly disagree, Don't know/Not sure) Used by CDC Adult Tobacco Survey; COMMIT evaluation
Comments	Recognition of the health benefits of cessation may be an important determinant of quit attempts, and an early indicator of the effects of health education efforts.
Measure 2: Youth	"What was the main reason you decided to stop smoking?" (SELECT ONE RESPONSE ONLY) a. I have never smoked cigarettes b. I have not stopped smoking c. To improve my health d. To save money e. Because my family does not like it f. Because my friends don't like it g. Other
Source	GYTS, 2007
Validity	Face validity.
Variations	
Comments	Recognition of the health benefits of cessation may be an important determinant of quit attempts, and an early indicator of the effects of health education efforts.
Measure 3: Awareness of Specific Resources	"Are you aware of assistance that might be available to help you quit smoking, such as telephone quitlines, local health clinic services, and…?" (Yes, No)
Source	GATS, 2007
Validity	Face validity.
Variations	"In the last month, that is, since [date], have you noticed any advertisements for stop-smoking medications?" (Yes, No) Used by the International Tobacco Control Policy Evaluation Survey
Comments	The "?" refers to (and should be replaced by) locally specific help promoted in the specific communication campaign. Awareness of smoking cessation resources increases the likelihood that smokers will make quit attempts. Information on the reach of interventions enables states to assess and improve the delivery of available resources.

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

Construct	(d) Tobacco Industry – Awareness and Knowledge
Measure 1: Adults	"I am going to read you some statements about tobacco companies. Please tell me if you strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with each of the following statements: • Tobacco companies should be allowed to advertise and promote cigarettes as they please. • Tobacco products should be more tightly regulated. • Tobacco companies can be trusted to tell the truth about the dangers of their products. • Tobacco companies should take responsibility for the harm caused by smoking. • Tobacco companies have tried to convince the public that there is little or no health risk from secondhand smoke. • The government should do more to tackle the harm done by smoking. • The government doesn't really care about people smoking because it makes so much money from tobacco taxes."
Source	The ITC Project, 2007
Validity	Face validity.
Variations	
Comments	Recommended for use as single items when space is limited: • Tobacco companies can be trusted to tell the truth about the dangers of their products. • Tobacco companies have tried to convince the public that there is little or no health risk from secondhand smoke.
Measure 2: Youth	"People have different views about the issue of smoking and cigarette companies. How much do you agree or disagree with the each of the following (strongly agree, agree, disagree, strongly disagree): (RANDOMIZE ORDER) • Cigarette companies should have the same right to sell cigarettes as other companies have to sell their products. Would you say you… strongly agree, agree, disagree or strongly disagree? • Cigarette companies lie. • Cigarette companies deny that cigarettes cause cancer and other harmful diseases. • Cigarette companies deny that cigarettes are addictive. • Cigarette companies have done some really bad things. • Cigarette companies try to cover-up all the bad things they have done. • I would not work for a cigarette company • The people who run cigarette companies know what they are doing is wrong • No other companies act as badly as cigarette companies. • I would like to see cigarette companies go out of business. • Cigarette companies target teens to replace smokers who die • Cigarette companies get too much blame for young people smoking. • Anti-smoking advertisements are no more honest than cigarette ads. • Cigarette companies should have the same right to make money as any other type of company. • The government should let companies sell whatever they want. • Cigarette companies try to get young people to start smoking • Cigarette companies target minority groups."

Source	Legacy Media Tracking Survey (LMTS, (http://americanlegacy.org)
Validity	Established validity. Hersey *et al.*, 2005; Thrasher & Jackson, 2006.
Variations	
Comments	A five item scale measuring perceptions of the tobacco industry, based on "Cigarette companies lie," "Cigarette companies try to get young people to start smoking," "I would like to see cigarette companies go out of business," "I would not work for a cigarette company," and a fifth item "How much do you like cigarette companies?" (5 point scale: I like them a lot [1] to I don't like them at all [5]) showed small but significant improvement following the introduction of anti-tobacco industry media campaigns in selected US states (Hersey *et al.*, 2005).

LMTS: Legacy Media Tracking Survey ; ITC: International Tobacco Control Policy Evaluation Study; GATS: Global Adults Tobacco Survey; GYTS: Global Youth Tobacco Survey; COMMIT: The Community Intervention Trial for Smoking Cessation

Table 5.45 Measures to Assess Population Level Awareness and Knowledge of Public Communication Campaign Messages

cessation, but the actual objectives are to cast the company in a sympathetic light, change adult opinions about the culpability of the company in promoting tobacco use, and reinforce policy maker opinions about the company as socially responsible. Pharmaceutical company campaigns may be more transparent than tobacco company campaigns; that is, pharmaceutical campaigns that appear to promote the use of a particular cessation medication may be attempting to do exactly that. Public communication campaign planners may want to respond directly to tobacco company campaigns by countering or exposing the main purpose of tobacco company messaging in their own public communication campaigns, and may want to build on, reinforce, or avoid direct competition with pharmaceutical ads promoting proven cessation strategies. While tobacco and pharmaceutical companies have vastly greater resources to invest in marketing and communication, frequently public health programmes and governmental and nongovernmental organisations have access to channels that are off limits to tobacco companies (e.g. television and radio). Public communication campaign planners should avoid direct competition with tobacco companies, and instead utilize tools and strategies that give public campaigns the advantage (e.g. electronic media not available to tobacco companies and community action that exposes the human face of the tobacco tragedy).

To monitor the larger information environment, public communication campaign planners and evaluators may include indicators of awareness of and receptivity to tobacco and pharmaceutical company advertising on the same population surveys used to monitor campaign indicators (Farrelly *et al.*, 2002, 2003b), as well as attitudes toward, salience of and perceptions, beliefs and behaviours associated with exposure to the tobacco or pharmaceutical company campaign. Reach and frequency of these campaigns may be gleaned, imperfectly, by identifying print advertising and calculating impressions, and monitoring the airwaves for the appearance of ads and calculating exposure based on observations of placement. Tobacco and pharmaceutical company public relations and media advocacy efforts may be monitored through

the same news media tracking systems and content analysis undertaken to monitor implementation and outcomes of the public communication campaign (National Cancer Institute, 2005, 2006). While exact quantification of tobacco and pharmaceutical company campaigns may be unnecessary (or impossible) to obtain, a realistic understanding of the content and purpose of these competing messages is essential to crafting a meaningful and relevant public communication campaign that will be effective in a cluttered and contentious information environment.

Summary and recommendations

This section provides a framework for developing, implementing, and evaluating public communication campaigns. These multicomponent interventions seek to improve awareness and knowledge of tobacco-related issues with the intention of promoting individual behaviour change and support for and progress toward policy and social change. The purpose of evaluating these campaigns is to inform the development of effective campaigns, to identify and correct problems while the campaign is in process, and document the public health impact of the campaign. Core methods include testing campaign messages during the design phase, monitoring the reach of the campaign during implementation, and assessing core constructs, including awareness, knowledge, attitudes, beliefs, and support for policies and tobacco-related behaviour change. The measures described here, like the campaigns themselves, need to be customized to the specific content, purpose, and message of the communication effort being implemented.

Regardless of the results of the public communication campaign (and particularly if it failed to show results), evaluations should be made publicly available. A system to collect and document campaign results would enhance our understanding both of how public communication campaigns work and how to make them better.

5.7 Measures to assess the effectiveness of tobacco cessation interventions

The importance of encouraging smokers to quit completely is reflected in the actions outlined in Article 14 of the WHO FCTC (Figure 5.35), and has also been recognized by the World Bank as necessary in order to reduce tobacco related deaths in the next half-century (Jha & Chaloupka, 1999). Tobacco control interventions described here, and elsewhere in this Handbook, are expected to motivate smokers to make quit attempts. However, some smokers, especially those who are nicotine dependent (see Section 3.3) will need support in order to be able to stop successfully; that support is the main subject of this section.

In many countries, even though the majority of smokers want to stop smoking and many try to do so, they have difficulty succeeding. For example, in the UK, where there is a long established tobacco control movement, the natural population cessation rate is only about 1-2% per year. Smoking is a chronically relapsing condition and tobacco use has been recognized to be highly addictive (US Department of Health and Human Services, 1988; Royal College of Physicians, 2000). Tobacco dependence and withdrawal syndrome are classified as substance use disorders under the WHO International Classification of Diseases (WHO, 1992), and nicotine dependence and nicotine withdrawal are classified similarly by the American Psychiatric Association's Diagnostic and Statistical Manual (American Psychiatric Association, 1995).

1. Each Party shall develop and disseminate appropriate, comprehensive and integrated guidelines based on scientific evidence and best practices, taking into account national circumstances and priorities, and shall take effective measures to promote cessation of tobacco use and adequate treatment for tobacco dependence.

2. Towards this end, each Party shall endeavour to:

   (a) design and implement effective programmes aimed at promoting the cessation of tobacco use, in such locations as educational institutions, health care facilities, workplaces and sporting environments;

   (b) include diagnosis and treatment of tobacco dependence and counselling services on cessation of tobacco use in national health and education programmes, plans and strategies, with the participation of health workers, community workers and social workers as appropriate;

   (c) establish in health care facilities and rehabilitation centres programmes for diagnosing, counselling, preventing and treating tobacco dependence; and

   (d) collaborate with other Parties to facilitate accessibility and affordability for treatment of tobacco dependence including pharmaceutical products pursuant to Article 22. Such products and their constituents may include medicines, products used to administer medicines and diagnostics when appropriate.

WHO (2003)

Figure 5.35 WHO FCTC Article 14: ***Demand reduction measures concerning tobacco dependence and cessation***

Support for tobacco users trying to quit is incorporated in the range of tobacco control strategies, and complements the other approaches described in this Handbook. Implementing some of the interventions described in this section will need significant investments of time and money. It may be more appropriate for a country at an early stage of tackling the tobacco problem to focus on the strategies described in other sections (such as taxation and smoke-free policies), which will be less costly to implement. Strategies, such as smoke-free policies, also help to normalize non-smoking thereby providing an environment which motivates tobacco users to make attempts to quit. Nevertheless, some of the less intensive strategies described in this section can be promoted and implemented with ease. For countries which have implemented a comprehensive tobacco control strategy, the interventions described here become even more important.

In this section we refer to support for smokers when trying to stop as "tobacco cessation interventions." Tobacco cessation interventions are sometimes referred to as "treatment interventions," and for the purpose of this section, these terms will be used synonymously. Our definition of tobacco cessation interventions originates from Raw and colleagues (2002), who defined treatment interventions as including "…(*singly or in combination) behavioural and pharmacological interventions such as brief advice and counseling, intensive support, and administration of pharmaceuticals, that contribute to reducing or overcoming tobacco dependence in individuals and in the population as a whole*."

In many countries, tobacco cessation interventions are not widely available or integrated into healthcare systems. The availability and accessibility of pharmacological medications for smoking cessation also varies from country to country (Jha & Chaloupka, 1999). Tobacco dependence cessation interventions, in most countries, are often not as available as treatment for other addictions, such as illicit drugs and alcohol, suggesting that the addictive nature of tobacco use has not been adequately recognized and addressed.

This section provides protocols for measuring the existence and effectiveness of different forms of tobacco cessation interventions based on measures outlined in Article 14 of the WHO FCTC (Figure 5.35), and following a proposed conceptual framework for the evaluation of tobacco cessation policies and interventions (Figure 5.36). Since Article 14 only provides a minimum standard, this section builds on the measures advocated. It is the view of this working group that cessation interventions and policies should be evidence-based.

This section mainly focuses on interventions aimed at adult smokers, as most of the research has been carried out on them. However, cessation interventions are also targeted at sub-groups of the adult population, such as pregnant smokers or smokers in disadvantaged groups. Adolescent smokers may also be the target of cessation interventions. Target group considerations are important and should be taken into account when developing protocols and carrying out research (Chesterman *et al.,* 2005).

Policy

Figure 5.37 sets out various cessation policies, including the infrastructure thought necessary to implement cessation policies and interventions (e.g. evidence-based guidelines for tobacco cessation policies and interventions). Some countries have adopted these as a first step towards implementing cessation policies. Guidelines have been developed and implemented, *inter alia*, in the USA, Europe, UK, Canada, Australia, and New Zealand (Fiore *et al*., 1996, 2000; Raw *et al,* 1998, 2002; West *et al,* 2000; US Department of Health and Human Services, 2008). In addition, Figure 5.37 lists the type of interventions that a country may deliver. Evidence-based cessation interventions range from less intensive interventions that can be delivered on a large scale, such as brief, opportunistic advice by healthcare professionals, to more intensive interventions delivered to smokers either individually or in groups by a trained healthcare professional. Government smoke-free policies are also relevant to

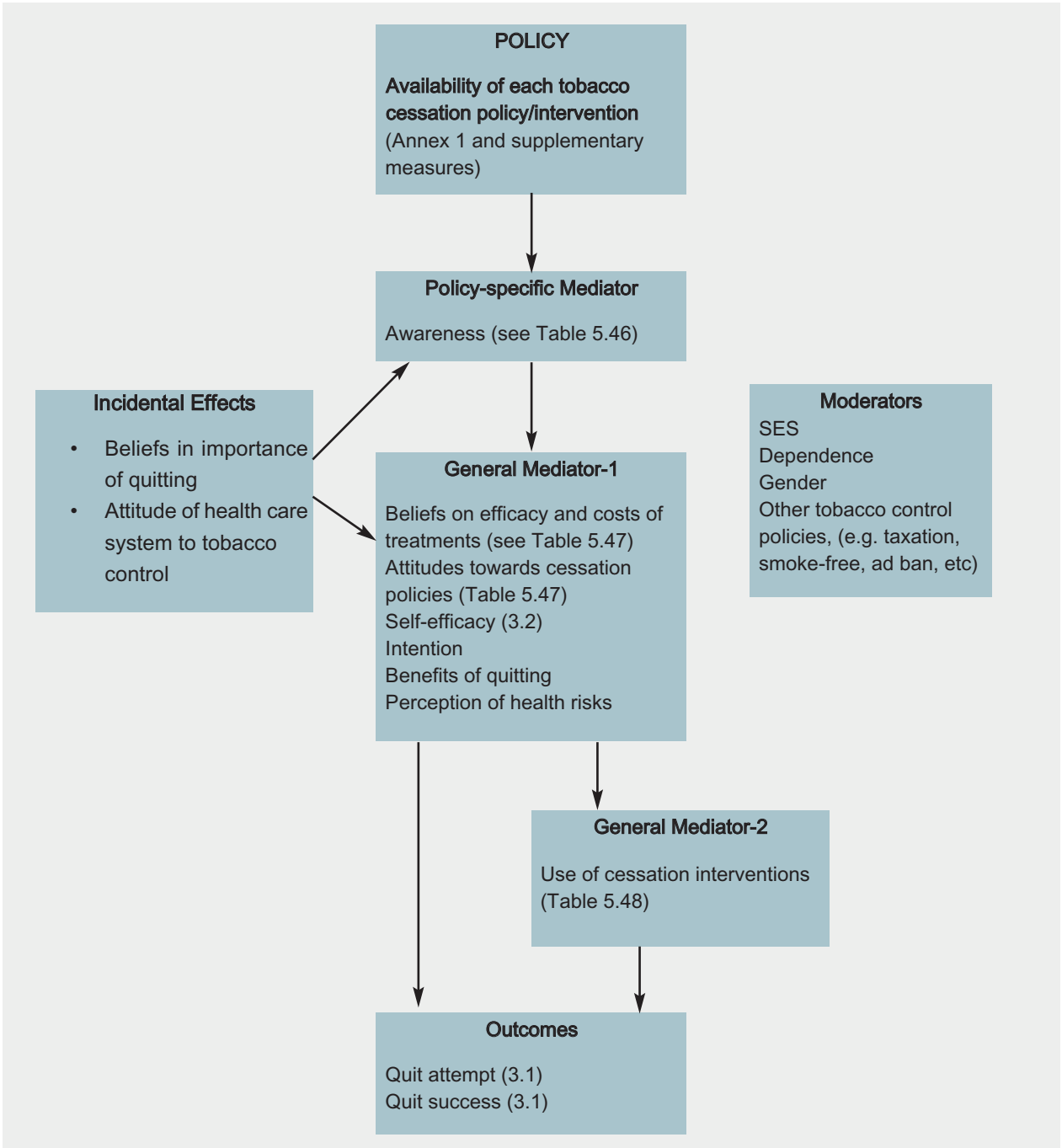


Figure 5.36 Conceptual framework for the evaluation of tobacco cessation policies and interventions

Numbers in parentheses indicate sections in the volume covering the argument

- Government policies & infrastructure on tobacco cessation
  - Availability & implementation of national evidence-based cessation guidelines
  - Existence of a smoking cessation coalition or partnership
  - Training for smoking cessation
  - Advertising/marketing of cessation interventions (e.g. helpline from government, nongovernmental or private (pharmaceutical) sources)
  - Development of a formal research programme for tobacco cessation
  - National quitline number advertised on packs/adverts
  - Reimbursement or level of funding or subsidy, for smoking cessation treatments including pharmacological interventions and mechanisms to provide this, eg through workplaces
  - Availability of tobacco cessation interventions:
    - Brief opportunistic advice being delivered routinely by doctors
    - A telephone helpline (preferably freephone) for smokers & promotion of it
    - Smoking cessation treatment services
    - Stop smoking medications – over the counter, prescription, give aways, approval of new medications, marketing rules
    - Quit and win contests
    - No smoking days
    - Mass media quit campaigns (see Section 5.6)
    - New technologies such as internet smoking cessation support and automated email messaging
    - Intensive cessation services delivered face to face either individually or in groups, consisting of behavioural interventions with pharmacological support

Figure 5.37 Government tobacco cessation policies and infrastructure for tobacco cessation

this section, but are covered in Section 5.2.

Figure 5.37 does not give an exhaustive list of the types of cessation policies or interventions that countries can offer, but outlines the relevant ones that a country is likely to adopt to satisfy Article 14 of the WHO FCTC, most of which have proven effectiveness. Early studies of the newer technologies, such as text messaging on on-line smoking cessation support, are promising.

Efficacy of cessation interventions

A summary of the efficacy of most of the interventions listed in Figure 5.37 can be found in 11 languages on the *Treatobacco* website (http://www.treatobacco.net). A few of the policies and types of infrastructure have not been evaluated, but as they are listed in Article 14 and are believed to be necessary to implement smoking cessation interventions, they are retained in this section.

When evaluating population effectiveness and impact of cessation interventions, the two key factors to be considered are reach and efficacy (effect size). Generally, interventions which are low intensity are more likely to reach a greater number of smokers within a population than high intensity interventions, but will have smaller efficacy. Conversely, more intensive interventions are more effective and will provide a greater degree of contact between the smoker and

the provider than low intensity interventions, but will not reach as many people and, therefore, may not have a measurable population impact. However, these more intensive interventions are important, for example, to highly dependent smokers who are likely to need more intensive support to stop smoking successfully. Though more intensive interventions are expected to incur a higher financial cost, their greater efficacy still makes them more cost effective compared with other healthcare interventions (McNeill *et al.*, 2005).

When appraising evidence of efficacy, it is important to know how abstinence has been measured. In 2003, the Society of Research on Nicotine and Tobacco (SRNT) convened a series of workgroups in order to provide guidance on measures used in clinical trials of treatments for smoking cessation (SRNT Subcommittee on Biochemical Verification, 2002; Hughes *et al.,* 2003, 2004b). The papers emanating from these groups provide "gold standards" to which those working in the smoking cessation field should aspire. These include the use of biochemical samples to validate self-reports of abstinence, such as expired-air carbon monoxide (CO) or cotinine level (a metabolite of nicotine) at various stages of follow-up, with the optimum being six months or longer. The rate of relapse after six months has been estimated enabling some assessment of quitting permanently from those followed-up for six months.

In some countries, and with some types of intervention (i.e. high reach, low efficacy), biochemical validation of quit rates at six and twelve months post-treatment may be cost-prohibitive with low compliance. We suggest as a minimum, data be collected on point prevalence at the end of treatment and six months later, with a random sample (if not all) of self-reported quitters being biochemically validated (see Section 3.2).

Assessing the existence and extent of implementation of tobacco cessation policies in a country

A simple questionnaire administered to policy makers, commissioners, or auditors will enable assessment of the availability of cessation policies, guidelines, interventions, and training within a country. Several tools have been developed to do this. A WHO Assessment Tool covers the availability of cessation services for tobacco dependence under five domains: infrastructure, support for treatment, intervention and treatment, healthcare providers, and healthcare users (Anderson, 2006). [In February 2008, WHO published the Report on the Global Tobacco Epidemic, available online at http://www.who.int/tobacco/mpower/en/, which outlines help offered to quit tobacco use by country in the world conveniently summarized in continent-specific spreadsheets]. Joossens and Raw (2006) recently developed a new Tobacco Control Scale to measure country activity and included measures for assessing treatment. In this scale, treatment is given a maximum of 10 points (out of a maximum of 100 points given for all tobacco control measures), with a maximum of two points being allocated for a quitline, six points maximum for a national network of specialized smoking cessation experts or units offering individual or group support delivered by properly trained professionals, and a maximum of two points for reimbursement of medications. The Framework Convention Alliance (2007) has documented the availability of treatment with a number of questions about government policy, clinical guidelines, promotion of cessation treatments, and the availability of individual interventions, as well as accessibility of medication in participating countries.

There is no easy way of validating the responses to questionnaires seeking cessation services information. Ideally more than one policy maker/ regulator/programme manager should be required to complete the questionnaire and supporting evidence sought via documentation. A recent review of the array of availability of cessation interventions within England may also be helpful in determining the type of information that should be collected (McNeill *et al.,* 2005).

Cost data will be needed to measure cost-effectiveness of cessation policies. For each cessation policy or intervention, the costs both to the provider and

the smoker utilizing it should be assessed. Together with an assessment of the likely benefits, an estimate of cost-efficacy can then be made (Godfrey *et al.,* 2005).

Supplementary measures will be needed to assess the implementation of specific smoking cessation interventions, in order to understand data on smokers' usage and perception of these interventions. Examples of the types of data that can be collected for some common cessation interventions follow.

Supplementary measures needed to assess availability of specific cessation interventions:

a) Brief opportunistic advice by healthcare professionals

The key measure of interest is whether smokers recall receiving advice to quit smoking from healthcare professionals, and whether they report acting on this advice (see below). However, it can also be useful to supplement such data with an assessment of the proportion of healthcare professionals who report offering smoking cessation advice, as some smokers may not recall receiving advice, or deny receiving it. Surveys of healthcare professionals often demonstrate higher levels of reported intervening than is suggested in surveys of smokers, which may suggest that healthcare professionals overestimate their frequency of discussing interventions. Interpretation of these findings can be facilitated by qualitative research, such as the use of observational techniques to better understand the context within which brief interventions are given, if they are given at all, and why the advice may not be having the impact that is desired (e.g. if it is too brief) (Coleman *et al.,* 2004). It can also be useful to assess doctors' views of referring smokers for further support (McEwen *et al.,* 2005).

To be able to advise smokers to quit, healthcare professionals need to keep up-to-date records of the smoking status of all patients, and be aware of more intensive support that is available to which smokers can be referred as appropriate. Auditing notes about patients can help assess whether smoking status and interventions (such as advice to quit, prescriptions, referrals) are being recorded in a systematic and consistent way, and can assess the availability of reminder systems for healthcare professionals to intervene on smoking matters (Anderson & Jane-Llopis, 2004).

b) Telephone helplines

For countries running telephone stop smoking helplines, monitoring is needed to answer questions about their purpose, target audiences, reach, cost, and effectiveness. The different purposes that telephone helplines can serve need to be identified (Centers for Disease Control and Prevention, 2004). The most common is to act as a first port-of-call for smokers seeking help (e.g. following a television or radio advertisement). Smokers contacting the helpline can then be given support either in the form of self-help materials, brief or intensive counseling, or they can be directed to other sources of information or support. The helpline may also be used proactively and involve multiple call-backs offering further support. If the helpline is used in conjunction with media campaigns, the evaluation of the helpline would need to be assessed alongside the evaluation of the media campaign. In this case, the outcome measures for the helpline evaluation should directly link to its intended purpose in relation to the mass media. For example, if the purpose is to direct smokers where to go for further help, assessing whether information on effective treatments was given out (and subsequently used), is very different from an assessment of effectiveness if the purpose is to deliver a smoking cessation intervention (Centers for Disease Control and Prevention, 2004).

Alternatively, the purpose of the helpline may be to target specific groups, such as pregnant smokers. Basic demographic and tobacco use data (see below) are useful for assessing the ability and success of the helpline in reaching its stated target groups. As some target groups are difficult to reach, progress can be compared between a newly set-up helpline and one that is well-established and strives to reach similar target groups (Centers for Disease Control and Prevention, 2004).

Collecting data consistently across different helplines will aid in such comparisons.

The CDC (2004) also recommends collecting process evaluation data, for example, call volumes, how many callers get different types of service, how many callers get through to a live counselor, etc. Understanding how the service is utilized and factors affecting the caller's choice of service will help make sense of effectiveness and cost data. Knowing how callers heard about the helpline will be important to be able to assess which channels of advertising are most cost-effective. Caller satisfaction is also useful to assess (e.g. are callers getting the service they were expecting? do they receive the materials they were told they would? how long did they wait to speak to a counselor?). Caller satisfaction can also be assessed by asking open ended questions of a random sample of callers.

c) Stop smoking medications

It is worthwhile trying to obtain sales data for stop smoking medications in countries. Often these data will need to be obtained from market research companies (e.g. aggregate sales data on pharmaceuticals; companies which collect aggregate sales data on pharmaceutical sales, such as IMS Global Services, AC Nielson, and IRI). Alternatively, the pharmaceutical manufacturers might be able to get permission to share sales data from the market research companies. The limitations of using commercial sales databases are discussed in Section 3.5. Sales data from pharmacies might also be available. Sales data can be evaluated to assess the impact of changes in policies or accessibility (West *et al.*, 2005). Government data can also be sought on medication subsidies or prescription script receipts.

d) No Smoking Days

In addition to cost and target group, message type and media penetration can also be monitored for No Smoking Days.

e) Quit and Win contests

Similar process indicators to those referred to above, for No Smoking Days, can also be monitored here.

f) Intensive cessation services

It would be helpful to know how many services exist in a particular country and any monitoring data that is routinely collected. A comprehensive evaluation of a national network of smoking cessation services was recently carried out in England (Raw *et al.*, 2005). This study included an evaluation of monitoring data collected by the services to evaluate short- and long-term outcomes (Ferguson *et al.,* 2005; Judge *et al.,* 2005). Guidance exists on the monitoring data most useful to capture on a routine basis (McNeill *et al.,* 2005; West, 2005b).

In addition, surveys (qualitative and quantitative) can be carried out with healthcare professionals dedicated to giving specialist smoking cessation advice and support. Such surveys were recently conducted as part of the national evaluation of smoking cessation services in England (Bauld *et al.,* 2005; Coleman *et al.,* 2005; Pound *et al.,* 2005). These surveys enable an assessment of the perceived barriers to giving adequate advice and support to smokers.

Policy specific mediators (proximal measures)

Smokers need to be aware of the availability of cessation interventions before they can access them. Questions can therefore be asked about awareness of support that is available to help smokers quit and whether they are aware that they can get financial support for treatment or free cessation treatment (see Tables 5.46a and 5.46b).

Consumer surveys with smokers and recent ex-smokers (usually defined as smokers who have stopped within the last year) can assess awareness for different types of smoking cessation policies and interventions. It may also be important to ask how consumers hear about different types of interventions to help assess the most appropriate communication routes to profile these interventions. If appropriate, it might also be useful to examine these results by target group (e.g. pregnant women). It is also possible, although more resource intensive, to carry out separate

Construct	(a) Awareness of Tobacco Cessation Interventions
Measure 1	"Are you aware of assistance that might be available to help you quit smoking, such as telephone quitlines, local health clinic services?" (Yes, No, Don't know)
Source	US Adult Tobacco Use Survey from CDC (Starr *et al.*, 2005)
Validity	Unknown - face validity.
Variation	Could be expanded to include a comprehensive array of culturally and country-specific tobacco cessation interventions.
Comments	Researchers might want to include a follow-up question to assess which sources of cessation services individuals are aware of (e.g. "If Yes, what is available to help you quit?").
Construct	**(b) Awareness of Tobacco Cessation Intervention Reimbursement**
Measure 1	"Does any of your health insurance cover treatment to quit smoking cigarettes or to stop using other tobacco products?" (Yes, No)
Source	American Smoking and Health Survey from CDC (Starr *et al.*,2005)
Variation	Could be expanded to assess awareness of the specific types of cessation interventions covered (e.g. counseling, medication), rather than coverage in general.
Validity	Unknown - face validity.
Comments	Should be adapted to other countries where treatment might be financed by sources other than insurance. This measure isn't relevant to individuals who do not have insurance.
Construct	**(c) Awareness of Tobacco Cessation Intervention Medications**
Measure 1	"Have you heard about medications to help people stop smoking, such as nicotine replacement therapies like nicotine gum or the patch, or pills such as Zyban?" (Yes, No)
Source	The ITC Project, 2007
Variation	Include whatever medications are relevant for the country being surveyed.
Validity	Unknown - face validity.
Comments	Probably do not want to ask in some countries where awareness is ubiquitous.
Measure 2	"In the last month have you noticed any advertisements for stop-smoking medications?" (Yes, No)
Source	The ITC Project, 2007

Table 5.46 Population-Level Survey Measures of Awareness of Cessation Interventions, Reimbursement, Medications, and No Smoking Days

Variation	This could be expanded to include advertisements for other tobacco cessation interventions. Time reference should be specific to the policy implementation time-line.
Validity	Unknown - face validity.
Comments	Could be adapted for different tobacco cessation interventions.
Construct	**(d) Awareness of No Smoking Days**
Measure 1	"Some months ago, there was an organized day about smoking. Do you remember what it was called?" (Yes, No)
Source	Owen & Youdan, 2006
Variation	Adapt or tailor according to how the day is referred to in a country.
Validity	Unknown - face validity.
Comments	
Measure 2	"No Smoking Day was held on [date]. Do you remember it?" (Yes, No)
Source	Owen & Youdan, 2006
Variation	Adapt or tailor according to how the day is referred to in a country.
Validity	Unknown – face validity.
Comments	
The ITC project: The International Tobacco Control Policy Evaluation Survey	

Table 5.46 Population-Level Survey Measures of Awareness of Cessation Interventions, Reimbursement, Medications, and No Smoking Days

surveys for smokers and ex-smokers for each individual intervention allowing for more comprehensive data to be assessed.

Examples of questions that can be used in surveys of smokers and recent ex-smokers to assess awareness of specific smoking cessation interventions are shown in Tables 5.46c and 5.46d (in this case smoking cessation medications and national No Smoking Days). Countries at an early stage of the tobacco epidemic may consider asking smokers and recent ex-smokers whether they are aware healthcare professionals can offer advice or support to stop smoking. It might be appropriate to separate questions asking about advice from doctors from questions about advice from other healthcare professionals, depending on which professional groups are being targeted to offer assistance within a country.

General mediators (intermediate measures)

It can be important to measure smokers' attitudes towards government cessation policies and interventions. Such questions can shed light on whether tobacco

users perceive their tobacco use as an addiction, in a similar way to other addictions, and whether they therefore believe it is appropriate for governments to be offering support in stopping.

Questions assessing the proportion of smokers who believe that specified cessation methods will help them to quit, can be useful for assessing whether smokers are distinguishing between unproven and proven methods and recognize the importance of seeking help with quit attempts. Beliefs about whether cessation support should be free to smokers also reflects whether smokers believe that getting help can increase the likelihood of their quit attempt being successful or whether really only willpower is needed. Examples of these types of questions are given in Table 5.47a.

Questions can also be asked about barriers to seeking help with stopping. Table 5.47b gives an example of a question assessing perceived barriers to using smoking cessation medications.

Measuring beliefs about the role of nicotine (Table 5.47c) will also help to elucidate whether smokers understand that they are or might be dependent on nicotine. Such questions will help to identify whether they are distinguishing between habit and addiction, which will also help to understand their responses to questions on seeking help in stopping. Questions about beliefs on nicotine will also help clarify their understanding of how nicotine replacement medications might help them to stop (Siahpush *et al.,* 2006).

Specific questions can be asked about individual cessation interventions that smokers are aware of, and for each one their beliefs about usefulness and perceived efficacy. Table 5.47d shows such a question for No Smoking Days.

General mediators (distal measures)

A general question can be asked to assess which cessation interventions, if any, smokers and ex-smokers used when trying to stop tobacco use recently. The time interval period over which smokers/ex-smokers should be asked to recall interventions needs consideration. Smokers have been shown to forget quit attempts, particularly shorter ones (West, 2006), so if the period is too long this is likely to result in increased forgetting. However, having a period which is too short will increase the likelihood of missing some events of interest. An alternative way of asking questions about intervention use is to link a quit attempt (e.g. the most recent quit attempt) with the support used, rather than asking what methods have been used over a time period. This makes it easier to ascertain which methods most likely contributed to quit attempts and success, but will miss some attempts to quit. Probably a combination of the different types of questions is needed. An example of a question which can be adapted to test use of interventions either over a time period or during a recent quit attempt is given in Table 5.48a.

As well as generic questions, smokers and recent ex-smokers can also be asked further details about specific cessation interventions such as how they were accessed or how they were used. Some examples of these are covered in the sections below. Questions can also be asked about correct use or compliance as well as any perceived impact.

a). Advice by healthcare professionals

Surveys of smokers (and recent ex-smokers) can assess whether they have visited healthcare professionals, whether they recall being asked about their smoking and their motivation to quit, and whether they recall receiving advice to quit or support from healthcare professionals and how they acted on the advice (see Table 5.48b). They can also be asked whether there were any follow-ups offered or arranged by their healthcare professionals.

b). Stop smoking helplines

Evaluating a quitline can involve taking a random sample of callers and following them up to see how many quit after a period of time, for example six months (Centers for Disease Control and Prevention, 2004). Though this method is relatively straightforward to carry out, it cannot

Construct	(a) Beliefs About the Benefits of Tobacco Cessation Interventions
Measure 1	"Which of the following cessation interventions do you think would help you to quit?: a. Call a quitline b. See a physician c. Join a cessation programme d. Use a nicotine patch, gum, nasal spray, inhaler, lozenge, or tablet e. Use a prescription pill, such as Zyban, Bupropion or Wellbutrin f. Use an internet smoking cessation programme g. Quit with a friend, relative, or acquaintance h. Other method i. Quit on your own"
Source	Modified from CDC (Starr *et al.,* 2005)
Variation	Can be modified to assess any culturally relevant or country-specific cessation methods, either evidence-based or non-evidence-based.
Validity	Unknown - face validity.
Comments	Assesses to what extent and which individuals recognize that cessation interventions can help them. Has not been widely used to date. There is no ranking of what would be most helpful.
Measure 2	"I'm going to read a list of statements about stop-smoking medications. Please tell me if you strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with each of the following statements:" a. If you decided you wanted to quit, stop-smoking medications would make it easier. b. If you decided you wanted to quit, you would be able to quit without stop-smoking medications.
Source	The ITC Project, 2007
Variation	This question should be asked specifically of certain medications (e.g. various Nicotine Replacement Therapy, Bupropion).
Validity	Unknown - face validity.
Comments	These questions could be expanded to include specific medications and other non-medication cessation interventions.
Measure 3	"Proven therapies for treatment of tobacco dependence should be covered by health insurance plans." Do you: a. Strongly agree b. Agree c. Disagree d. Strongly disagree
Source	CDC (Starr *et al*., 2005)

Table 5.47 Population-Level Survey Measures of Beliefs about and Barriers to Using Tobacco Cessation Interventions, and Beliefs about Nicotine and No Smoking Days

Variation	Adapt for country-specific funding sources.
Validity	Unknown - face validity.
Comments	Only appropriate for countries with insurance. Could be modified to "free to smokers wanting to quit." The current item is somewhat poorly worded and may be confusing to respondents.
Construct	**(b) Beliefs About Barriers to Tobacco Cessation Interventions**
Measure	"I'm going to read a list of statements about stop-smoking medications. Please tell me if you strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with each of the following statements": a. Stop-smoking medications are too expensive b. You don't know enough about how to use stop-smoking medications properly c. Stop-smoking medications are hard to get d. Stop-smoking medications might harm your health
Source	The ITC Project, 2007
Variation	This question should be asked specifically of certain medications (e.g. various Nicotine Replacement Therapy, Bupropion)
Validity	Unknown - face validity.
Comments	These questions could be expanded to include specific medications and other non-medication cessation interventions. An item could be added to assess whether general costs of cessation represent a barrier (e.g. "Which of the following best describes your beliefs about the costs of quitting smoking: a) It's too expensive; b) It's expensive but if I wanted to I could afford it; and c) expense is not a problem").
Construct	**(c) Beliefs About Nicotine**
Measure	"Do you believe that the nicotine in cigarettes is the chemical that causes most of the cancers?"
Source	The ITC Project, 2007
Validity	Unknown - face validity.
Variation	This could be adapted to cover other diseases caused by smoking.
Comments	

Table 5.47 Population-Level Survey Measures of Beliefs about and Barriers to Using Tobacco Cessation Interventions, and Beliefs about Nicotine and No Smoking Days

Construct	(d) Beliefs About No Smoking Days
Measure	"Do you think No Smoking Day is a good/bad idea?" "What do you think the main purpose of No Smoking Day is?" "From what you remember, did you feel No Smoking Day was aimed at people like you, or not?" "I'd now like to talk about No Smoking Day in general. Did it make you feel more or less confident about stopping smoking or did it make no difference?"
Source	Owen & Youdan, 2006
Validity	Unknown - face validity.
Variation	These can be adapted to cover information on specific smoking cessation medications, if more than one type is available in a country, distributed during No Smoking Day.
Comments	
The ITC project: The International Tobacco Control Policy Evaluation Survey	

Table 5.47 Population-Level Survey Measures of Beliefs about and Barriers to Using Tobacco Cessation Interventions, and Beliefs about Nicotine and No Smoking Days

determine what proportion of the quitting is attributable to the helpline and what proportion would have quit without it; for this, a randomised controlled study would be needed which can have significant cost implications (Centers for Disease Control and Prevention, 2004).

CDC recommends that various issues be taken into account when reported quit rates are being assessed in the absence of a control group. These include: an exact description of how the callers contacting the helpline were selected for inclusion in the evaluation sample; a description of baseline caller characteristics, such as dependence and intention to quit, as this may affect quit success; a follow-up of a random sample of successes to ascertain long-term success, given loss to follow-up; and a calculation of success rate to assume those not followed-up relapsed to smoking. Appendix F of the CDC quitline report (Centers for Disease Control and Prevention, 2004) contains a recommended minimum data set for evaluating helplines, and Chapter 4 of the European Network of Quitlines Best Practice Guide provides similar information (European Network of Quitlines, 2004).

An alternative means of assessing the impact of reactive helplines on smokers' quitting behaviour at a national level, is to survey smokers (and recent ex-smokers) and ascertain whether they contacted a helpline, got through, and the impact of that intervention (Table 5.48a).

c) Stop smoking medications

Surveys of smokers (and recent ex-smokers) can assess whether they have accessed, purchased, and/or used stop smoking medications (Tables 5.48c and 5.48d). It is also important to ask how they used the medication (e.g. to cut down or to stop smoking altogether), for how long they used it, and whether they are still using the medication. Responses from population surveys to questions about accessing medications (either by purchasing

Construct	(a) Use of Tobacco Cessation Interventions
Measure	"Have you used any of the following to try and stop using tobacco?" (Yes, No) a. Counseling, including at a smoking cessation clinic? b. Nicotine replacement therapy? c. Other prescription medications, for example (FILL IN WHATEVER IS RELEVANT TO THE COUNTRY)? d. Traditional medicines, for example (FILL IN WHATEVER IS RELEVANT TO THE COUNTRY)? e. Acupuncture? f. Hypnosis? g. Quit line? h. Anything else? (Please specify:_____)
Source	GATS, 2007
Variation	Can be modified to assess any culturally relevant or country-specific cessation methods, either evidence-based or non-evidence-based. It can also be modified to specify which quit attempt is of interest (e.g. most recent, any quit attempts since a policy implementation).
Validity	Unknown – face validity.
Comments	Time scale can be varied to ask about ever used, used in last attempt, or used since policy implementation.
Construct	**(b) Receipt of a Tobacco Cessation Intervention from a Healthcare Professional**
Measure	"During any visit to a healthcare professional in the last 6 months, did you receive (Yes, No for each): a. Advice to stop smoking? b. Additional help or referral to another service to help you quit? c. Prescription for stop-smoking medication? d. Pamphlets or brochures on how to quit? e. Did they arrange a follow-up? f. Did not visit a healthcare professional in the last 6 months? During any visit to a doctor or healthcare provider in the past 12 months, did you receive advice to quit using tobacco?" (Yes, No)
Source	The ITC Project, 2007 (adapted to include follow-up); GATS, 2007
Variation	Can adapt for individual professionals (e.g. doctor, nurse, pharmacist).
Validity	Unknown – face validity.
Comments	Brief advice from a physician is efficacious.

Table 5.48 Population-Level Survey Measures of the Use of Tobacco Cessation Interventions (TCI), Receipt of TCI Information from Healthcare Professionals, Assessing the Use of Tobacco Cessation Medications, How Medications were Obtained, and Behaviour Change on No Smoking Days

Construct	(c) Use of Tobacco Cessation Medication
Measure	”Have you used any stop-smoking medication?” (Yes, No, Can’t remember) ”In the last 6 months – since [6 month anchor] – have you used any stop-smoking medication?” (Yes, No, Can’t remember) ”In the last 6 months, which medication or medications did you use (do not prompt)?” (require type not brand name, can select more than one) ”The last time you used a stop-smoking medication, did you use more than 1 product at the same time?” (Yes, No) ”Which medications did you use at the same time?” ”For how long did you use the medication?”
Source	The ITC Project, 2007
Variation	The time scale can be adjusted to assess, all medication use, most recent use, or use since the policy implementation.
Validity	Unknown - face validity.
Comments	Could supplement or replace with pharmaceutical sales data.
Construct	**(d) Access Tobacco Cessation Medication**
Measure	”How did you get [medication from previous answer]?” (By prescription, Over the counter/off the shelf, From a friend) When you used [medications from previous answer], did you pay full price, get a discount, or get it free?
Source	The ITC Project, 2007
Variation	These may need to be changed to be country specific.
Validity	Unknown - face validity.
Comments	Could supplement or replace with prescription or pharmacy data.

Table 5.48 Population-Level Survey Measures of the Use of Tobacco Cessation Interventions (TCI), Receipt of TCI Information from Healthcare Professionals, Assessing the Use of Tobacco Cessation Medications, How Medications were Obtained, and Behaviour Change on No Smoking Days

or through a healthcare professional) can be compared with sales data and medication subsidies or pre-scription receipts.

d). No smoking days

Examples of questions used in annual surveys of the UK No Smoking Day are given in Table 5.48e.

e). Intensive cessation services

User satisfaction surveys can also be used, if appropriate, to increase understanding of why and how cessation services have a particular impact.

Summary and recommendations

Article 14 of the WHO FCTC obligates ratifying nations to adopt policies that promote access to evidence-based tobacco cessation interventions. Such interventions range from less intensive efforts, such as brief, opportunistic

Construct	(e) Behaviour Change on No Smoking Days
Measure	"Did you yourself attempt to give up or cut down your smoking on No Smoking Day?" (Yes, No) (for those who answer "No," ask why not) For those who say yes: "Did you….. a. Give up for the whole day? b. Give up smoking for part of the day? c. Cut down your number of cigarettes on that day? d. Or did you find you just couldn't cut your smoking? e. Can't remember?" For those who did stop or reduce, including on the Day itself, for how long did you manage to stop or reduce your smoking? How long did you intend to stop or reduce smoking? Why did you want to reduce or stop smoking on No Smoking Day?
Source	Owen & Youdan, 2006
Validity	Unknown - face validity.
Variation	These can be adapted to similar days in other countries.

GATS: Global Adult Tobacco Survey
The ITC project: The International Tobacco Control Policy Evaluation Study

Table 5.48 Population-Level Survey Measures of the Use of Tobacco Cessation Interventions (TCI), Receipt of TCI Information from Healthcare Professionals, Assessing the Use of Tobacco Cessation Medications, How Medications were Obtained, and Behaviour Change on No Smoking Days

advice by healthcare professionals, to more intensive efforts delivered to tobacco users either individually or in groups by trained healthcare professionals. Core constructs for evaluating access to tobacco cessation interventions include: proximal variables, such as awareness of cessation interventions; intermediate variables, such as specific beliefs and attitudes about different cessation interventions; and distal variables reflecting the utilisation of different cessation interventions.

The effects of policies facilitating access to tobacco cessation interventions can be assessed through self-report using standardised surveys of current and former tobacco users, and by reviewing records that document trends in utilisation of tobacco cessation interventions (e.g. calls to a helpline, sales of stop smoking medications). Measures described here are useful exemplars of how to assess utilization of cessation services. Evaluations of the effects of policies to promote access to cessation interventions should preferably include a longitudinal design, which assesses the relationship between the utilization of cessation treatments by current and former tobacco users and tobacco use behaviors.

Summary

Background

In the 20th century, cigarette smoking caused an estimated 100 million deaths worldwide (Gajalakshmi *et al.*, 2000). Most of these deaths were in high-resource countries where cigarette smoking first became popular in the 1920s to 1940s. This resulted in an epidemic of smoking-induced cancer, heart disease and chronic obstructive pulmonary disease (COPD) deaths. Cigarette smoking is not only the most prevalent form of tobacco use, it is also particularly harmful, killing one of two long-term users, half of them (one in four users) in middle age. In 2000, smoking was responsible for approximately 4.83 million deaths in people 30 years of age and older, evenly divided between high- and low-resource countries (Ezzati & Lopez, 2003), with lung cancer accounting for 0.52 and 0.33 million deaths, respectively (Ezzati & Lopez, 2004). If current mortality trends continue, it will cause some 10 million deaths each year by 2030, with around 70% in low-resource countries (Peto & Lopez, 2001). If present usage patterns persist, smoking will cause approximately 1 000 000 000 deaths this century, a tenfold increase over the previous century (Peto & Lopez, 2001). Most of these expected deaths could be averted if we rapidly institute effective programmes to both discourage tobacco use and to assist those addicted to tobacco to quit (IARC, 2007a).

Tobacco is a plant containing the psychoactive and addictive drug nicotine. Although nicotine is the main psychoactive ingredient of tobacco and the source of its addictiveness, it is otherwise a minor contributor to the harm (Benowitz, 1998). Most of the harm is due to other constituents in tobacco, particularly in tobacco smoke (IARC, 2004). The harms from tobacco mainly stem from long-term use, which the addictive nature of the product promotes.

Across its long history, tobacco has been processed and consumed in a wide variety of ways. The two main forms of use are smoking combusted tobacco, and taking unburned tobacco into the mouth or the nose (smokeless use). Over the 20th century, the use of cigarettes, primarily factory-made cigarettes, dominated tobacco markets in nearly all countries. Cigarettes have also been the focus of most tobacco research. The use of other smoked tobacco products is now of only minor importance, except in some areas, particularly the Indian subcontinent, where the use of bidis prevails. All forms of smoked tobacco are very harmful to health (IARC, 2004), and attempts to create less-toxic versions of these products have generally failed, largely because they have been unacceptable to consumers. Smokeless tobacco, which is generally less harmful than smoked tobacco because it does not involve inhaling smoke, but still carcinogenic to the oral cavity and pancreas (IARC, 2007b), is not used in many parts of the world, but it is common in some areas and its use is significant and increasing in some countries (e.g. Sweden; Foulds *et al.*, 2003). With some forms of smokeless tobacco there has been success in reducing toxins while maintaining consumer acceptability (Broadstock, 2007). Non-cigarette tobacco use is underresearched in comparison to cigarette use.

In recognition of the threat that tobacco use poses to global public health, in May 2003, the member countries of the WHO adopted the Framework Convention on Tobacco Control (WHO FCTC), the first international treaty devoted to improving public health by restraining tobacco promotion and use (WHO, 2003).

Scientific evidence plays a central role in the WHO FCTC. Its Foreword describes the WHO FCTC as "an evidence-based treaty that reaffirms the right of all people to the highest standard of health" (WHO, 2003). The preamble to the FCTC states that adopting nations are "determined to promote measures of tobacco control based on current and relevant scientific,

technical, and economic considerations" (WHO, 2003). To achieve its objective, the WHO FCTC calls for a comprehensive range of policies, defined for the purposes of this Handbook as the enabling mechanisms that allow particular rules, regulations and programmes to operate (in other words, frameworks that allow instruments to be implemented). The key articles of the Convention relevant to this Handbook are:

Article 6
Price and tax measures to reduce the demand for tobacco

Article 8
Protection from exposure to tobacco smoke

Article 9
Regulation of the contents of tobacco products

Article 10
Regulation of tobacco product disclosures

Article 11
Packaging and labelling of tobacco products

Article 12
Education, communication, training and public awareness

Article 13
Tobacco advertising, promotion and sponsorship

Article 14
Demand reduction measures concerning tobacco dependence and cessation

Article 15
Illicit trade in tobacco products

Article 16
Sales to and by minors

Article 17
Provision of support for economically viable alternative activities

Article 20
Research, surveillance and exchange of information

Article 22
Cooperation in the scientific, technical, and legal fields and provision of related expertise

The WHO FCTC is a seminal event in global health. Scientific evidence has demonstrated the enormous health harms of tobacco use. Scientific evidence as to the effectiveness of potential interventions formed the basis for the selection of the policies that are included in the WHO FCTC. However, whether the WHO FCTC is to fulfill its objective of reducing the devastation of the tobacco epidemic will depend on how effectively countries formulate and implement these policies. Moreover, history has shown us that the tobacco industry will adapt and work to circumvent even the strongest policies, so governments will also need to be ready to evolve and change their policies in order to ensure they achieve their goals. Good public health practice demands ongoing evaluation research as critical to informing the implementation and dissemination of established policy instruments as well as to aid in the subsequent evolution of new policy-related interventions.

Overview

This Handbook is concerned with the articulation of a framework and methods for conducting tobacco control policy evaluation, and not with an evaluation of a body of research in itself. It also offers terminology to judge the quality of the evidence considered in such evaluations and to be applied by IARC in the future evaluation of specific tobacco control policy interventions. As a result, the WG's advice to the potential readers of the Handbook is largely about how to evaluate policy interventions in ways that we believe will best advance tobacco control. In addition to this advice to researchers and evaluators, a small number of recommendations directed at other audiences are made.

The goals of this Handbook are to move the field by:

a) developing a common framework and language for tobacco control policy evaluation;
b) reviewing the strengths of possible research designs;
c) using theory to derive core constructs to measure when doing evaluations of key tobacco control policies;
d) identifying measures of constructs, and
e) providing an assessment of the scope and quality of existing

data sources. Four broad questions guided the review of the scientific literature on the methods and measures of tobacco policy evaluation:

1. How do we determine the effects of a policy?
   What are the key features of the policy as implemented?
   Is there a common conceptual framework that can be applied to understand how policies work?
   How might different design features be used to reduce threats to internal validity?
2. What are the core constructs for understanding how and why a given policy works?
   Which of these are parts of general pathways, and which are specific to particular policies?
   What is the quality of the measures used to assess core constructs?
   Do these measures, as well as the constructs they presumably reflect, translate into different cultures and contexts?
3. What are potential moderator variables to consider when evaluating a given policy?
   What is the quality of the measures used to assess potential moderator variables?
4. What data sources exist that might be useful for evaluation?

How useful are these data sources for evaluation (i.e. completeness and quality)?
The WG acknowledged that in attempting to answer these questions, explicit considerations must be given to equity issues both within and between countries. This involves always asking the question: "What is needed to optimise the intervention for disadvantaged groups?" This may range from making sure a programme is available in disadvantaged areas, to ensuring that the wording and tone of communications is acceptable and comprehensible.

The Handbook outlines a framework that interested organisations, including governments, can utilise to measure the effectiveness of interventions aimed at implementing tobacco control policies that are currently being and will be adopted in the next several years in adherence to the WHO FCTC. It describes major steps we made to articulate a new and coherent framework for thinking about tobacco control interventions.

The WG came from diverse disciplines, with different theoretical traditions and methodological approaches. This necessitated ongoing work to standardise language. We realised that some terminology was designed for thinking about the problem from a different perspective to the one necessary for understanding the complexity of population health areas like tobacco control. There is a need for ongoing work to rethink our terminology to better fit a population health framework.

The Handbook is intended to be a resource for researchers interested in evaluating tobacco control policies, and others interested in evaluating interventions beyond merely auditing implementation. It should also be useful for policy and programme developers as it spells out the theoretical frameworks upon which the interventions are based, and provides explicit models of how they exert their effects.

Steps towards a framework for evaluation

The WG began by considering what outcomes to focus on. It concluded, insofar as the interventions under consideration related to tobacco use and not to the harmfulness of each unit of the product, that the focus should be on tobacco use behaviours as the main outcomes of interest. This meant that, for the most part, the WG did not consider disease or mortality outcomes.

The WG concluded that there is currently no coherent framework for thinking about the evaluation of tobacco control policies in the policy literature. The frameworks borrowed from other areas such as clinical medicine are not adequate to the needs of the policy field. Randomised clinical trials are neither necessary nor often practical to generate evidence of the effectiveness of tobacco control policies.

The WG concluded that policy evaluation should be conceptualised in a manner analogous to how epidemiologists approach the task of inferring conclusions about the causes of disease (US Department of Health, Education and Welfare, 1964; Hill, 1965). This

is a framework that encourages researchers to triangulate all the available evidence to help rule out alternative explanations of observed effects, rather than focus on attempting to draw conclusions only from individual studies or from meta-analyses of studies using the same study design.

In the same way that evidence-based medicine has been built from rigorous evaluation of treatment options, evidence-based public health must begin with building a database from rigorous evaluation of public health policies. Evaluation of the effectiveness of tobacco control policies at the population level has been limited by inadequate data sources, problems in measurement and poorly conceptualised evaluation designs. It has also been limited by a failure to look for and maximise the value of studies with individually limited designs by systematically reviewing the findings from the corpus of such studies to determine what they collectively add to knowledge. In isolation or even combined in meta-analyses of similar studies, they may have little to tell us, but when they are combined in ways that take account of different threats to the validity of attributing causality by study type, they can sometimes be used to make strong inferences about causality as well as potentially increasing our understanding of the conditions under which the interventions are most effective. The benefits of such an approach are not just with regard to increasing our understanding of the effects of the intervention, but it also improves our ability to understand individual cases. Explicit comparison with the corpus of existing knowledge allows individual evaluators to say more about the programmes they evaluate than the designs they have adopted would allow them to do if they treated their evaluations in isolation of the accumulated knowledge.

The question one usually asks about policy interventions is: "Under what conditions can the desired effects be optimised?", not whether the intervention can work. Translated to the individual case, the question becomes: "Is the intervention working here as well as it should?" To answer that question one must be concerned about the form of the intervention, the ways it is delivered (quality of implementation), and various characteristics of the populations it is addressing. This is a framework that sees evaluation as part of a process of continual improvement. It is also about determining the relative contribution of each intervention to the overall goal, and how this might be moderated by characteristics of the broader environment. It involves paying more attention to the articulation of theoretical mechanisms, and having study designs that facilitate the elaboration of causal mechanisms.

Good evaluation starts with an analysis of the problem. Thus, the need to build an understanding of the factors that are affecting or can affect tobacco use and how use relates to the harms. Mechanisms by which tobacco control interventions can act to reduce harm must also be considered. The WG identified four aspects that need to be considered in evaluating interventions designed to reduce the harms. First, one must consider whether the goal of the intervention is to change tobacco use, tobacco harmfulness, or both of these. Second, a theoretical model or set of models describing how the interventions are expected to achieve their intended effects must be developed. Third, possible incidental effects of a policy that may occur must be considered. Fourth, any change in the environment that could modify the impact of the intervention (particularly counter-actions of the tobacco industry) must be monitored, and evaluated where necessary.

The first three steps in determining how policies may achieve their effects require specification of a theory of how the policy is expected to work. As Kurt Lewin noted years ago (1935), "there is nothing as practical as a good theory." The WG concluded that researchers should consider the adoption of a common framework to help identify relevant theories and thus guide the selection of core constructs useful for evaluating how and under what conditions tobacco control policies work. The issues that are likely to be relevant are to be considered well in advance. A general framework for assessing how an intervention might work is illustrated in Figure 6.1. At the first level it specifies two levels of mediating variables between a policy intervention and the outcomes: those specific to the policy, and those variables that are part of more

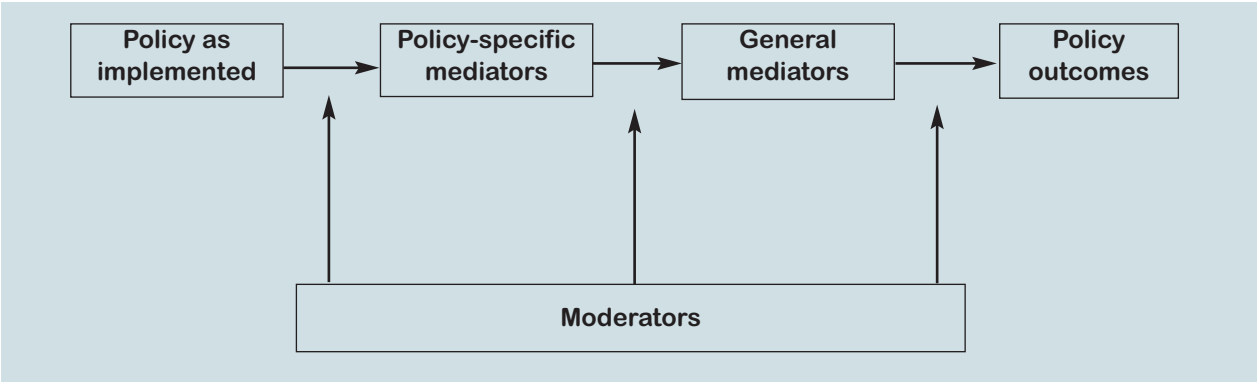


Figure 6.1 A generalised model of mediation making allowance for moderator effects

general pathways. It also accepts that various other factors (moderators) might affect the size of the effect.

There are only two main types of causal chain one needs to consider: the pathway from policies to tobacco use, and the pathway from tobacco products to levels of exposure to toxic substances and to the harms that result. Consideration of pathways may lead to the subdivision of a policy area into classes of interventions that share common pathways.

Understanding the mechanisms by which interventions have their effects is important because: 1) it can provide strong evidence of the causal impact of a particular policy, especially when attempting to differentiate the effects of a specific intervention from other possible causes, including other tobacco policies; 2) it can be used to diagnose the problem in cases where intended effects did not occur, by identifying where in the causal pathway things went wrong; 3) it can help us understand why a policy does not have the intended effect for some groups, but does for others (i.e. clarify why moderation occurs); and 4) in specifying how a policy works, it may help identify alternative ways of achieving the desired effects. These understandings can facilitate the development of new, and hopefully improved, ways of targeting key pathways of influence, or of tailoring interventions to better reach more resistant or needy groups.

The model outlines the primary constructs involved in helping to explain the relationship between tobacco control policies and their effects on tobacco use behaviours. In a limited number of cases, primarily in some aspects of product regulation, there is an alternative main path to outcomes—through reduced delivery of toxic chemicals. This is spelled out most clearly in the section on product regulation.

It is particularly important to go beyond the specific intent of some policies to explore their more distal ramifications. For example, the goal of information and product labelling policies is improved dissemination of knowledge to the potential user of the product. However, it is of interest to see whether and how these policies actually translate into changes in tobacco use behaviours. It is also important to consider effects along different pathways to the intended means of action, as these might be important for analysis of society-wide effects; e.g. the generally neutral or positive effects on business of smoke-free policies.

Finally, there needs to be consideration of unexpected effects on other determinants of tobacco use. This consideration is more important in tobacco control than in most other areas of health because such effects may be deliberately influenced by the tobacco industry (Cummings *et al.,* 2002b). Hence, surveillance of tobacco industry practices is required. The approach taken can be facilitated by a theoretical understanding of the industry's profit motive and marketing

practices, as this can guide the selection of data that are most relevant in surveillance for counter-active effects.

The conceptual framework for behaviour change assumes that each policy directed at changing tobacco use ultimately has its influence on those behaviours through a specific causal chain of psychological events. Policy-specific mediators involve such things as awareness, policy-specific knowledge and reactions to specific elements of the intervention. For example, new graphic warning labels should increase salience and visibility of warnings, and perhaps foregoing of occasional cigarettes. The second set of general mediators are constructs taken from behavioural science that we know mediate effects of behaviour; that is, they are means by which changes in tobacco use may occur. They include attitudes, normative beliefs and intentions. Moderators—those things that change the magnitude of the effects of an intervention without necessarily being changed by the intervention—include socio-demographic factors (e.g. age, gender, socio-economic status, cultural background) and psychological factors that are either assumed to be stable or which the intervention is not designed to change (e.g. level of dependence). This framework provides a general guide for thinking about policies and their effects on a broad array of important psychosocial and behavioural variables, and for testing how differences in policy implementation relate to effectiveness.

The model for the effects of changes in tobacco products to health effects can similarly be articulated, although here the distinction may be more between constructs that are measured in the environment (e.g. physical characteristics of cigarettes) and those within the individual (e.g. exposures, health harms), and the challenges of demonstrating links between the two—for example, the failure of current measures of cigarette yield to relate to measures of exposure to those chemicals in smokers.

The WG set the task of using diagrams, or logic models, to spell out the main factors to consider for each policy area and how they interrelate when considering all these policies simultaneously; to see if this approach would help elucidate common constructs and measures that might apply across different policy domains. The logic models allowed the WG to readily compare the similarities and differences in the constructs and measures across policy domains, and of the differences of policy type within a broad policy domain. The models were deliberately kept simple in an effort to focus attention on key constructs.

Finally, a major challenge is in the identification and validation of appropriate measures. Measurement validity is a particular issue, with measures of constructs varying in their validity dependent on the purpose they are being used for. This is sometimes because measures of known bias are used for measuring constructs because no better measures exist, but the differential effects of that bias in different contexts are overlooked.

The general theoretical framework presented here should be applicable across socio-cultural contexts. Clarification of policy intervention effects and the moderation of these effects will often involve comparative research. However, the specific theoretical model, its associated constructs and the measurement of these constructs may differ in important ways across national, cultural, linguistic and social groups. Where this happens, caution must be exercised in making comparisons between such groups.

Section Summaries

General methods and common measures

The Handbook first discusses features of research design for evaluation studies and how those features can form the basis for stronger conclusions about the impact of policies. Other aspects discussed and deliberated on include measurement issues in the design and analysis of cross-cultural comparative research, as well as some of the methods currently recommended for attempting to resolve these issues.

The Importance of Design in the Evaluation of Tobacco Control Policies

Evaluating the outcomes of population-level tobacco control policy involves three interrelated questions:

(1) Does the policy have an impact? (causality); if so,
(2) Under what conditions? (moderation); and
(3) How (mediation)?

The choice of design elements will depend on which questions are considered to be a part of the evaluation effort.

It is important to ensure that the appropriate concepts are chosen and that for each, measures are identified that are suitable to answer the evaluation question.

This section describes key design elements of outcome evaluation studies and how each contributes to reducing or eliminating threats to the internal validity of a study. Internal validity determines the extent to which the results of the study can lead to a causal conclusion.

Evaluation efforts should be informed by knowledge of the nature of the policy being evaluated, and the goals of the evaluation study should be clearly stated. Evaluation planning should be guided by understanding what threats to internal validity may be present in the study of a given policy situation, and then adding design elements and other measures to reduce or eliminate those threats.

Knowledge of the mediational pathways that are theorised to explain how policy affects behaviour and environment (or environmental risk) should lead to an appropriate study design, the inclusion of appropriate constructs and measures, and the selection of analytic tools that are well-suited to estimating the causal impact of policies by providing an explanatory pathway and helping to eliminate alternative explanations. Logic models describe these pathways and help identify constructs to measure. Suggestions on specific measures for many of these constructs are provided in other sections of this Handbook.

An outcome evaluation study must, at a minimum, include one post-policy measurement. In general, the addition of one pre-policy measurement (even cross-sectional) using the same measures and sampling frame is a more powerful evaluation strategy for assessing change due to a policy. The inclusion of a single, non-random control from another population is considered less desirable. Additional post-policy measurements are useful to track the effects of a policy over time. The utility of longitudinal designs is strengthened if there are multiple data collections before and/or after policy implementation, as this allows more precise specification of effects—for example, taking into account temporal trends that were occurring before the implementation of the policy. The role of time series analysis on aggregate sales/consumption data to demonstrating the effects of price on consumption is a good example of the power of multiple measurements.

Both repeated cross-sectional and longitudinal (cohort) designs are useful for assessing the impact of a given policy. The use of cohort designs provides additional capability for tracking the impact of policies within individuals, allowing stronger tests of mediational pathways.

Addition of samples from other populations to either or both intervention and control arms also adds strength to the evaluation design, as does having varying levels of intensity of the intervention.

Similarly, parallel assessment of alternative explanations for observed changes in outcomes (e.g. possibly being due to other policies or industry counter-actions) adds strength over assessing these effects in separate studies.

The existence of studies with complementary strengths and weaknesses is particularly useful in triangulating the results of a corpus of evaluation studies to see if a consistent pattern emerges.

The use of probability sampling in an evaluation study increases its external validity—the extent to which the findings of a policy evaluation study can be generalised to making conclusions about the impact of the policy on the larger population.

At a broader level, the design of an evaluation study should be guided by knowledge of how prior evaluation studies in the same policy domain have been conducted. An analysis of the similarity or differences in policy impact across similar studies can yield powerful conclusions about the overall impact of a policy.

Developing and assessing comparable questions in cross-cultural survey research on tobacco

Evaluation of tobacco control policies and other population-level interventions often involves data collection efforts across diverse national, cultural, linguistic and social groups. Comparison across such groups is often necessary to clarify policy effects, how these effects happen and how effects might differ across populations. The literature discussed in this section suggests that these comparative studies should consider measurement equivalence issues in the following ways:

Research teams should include collaborators from the socio-cultural groups in which the study is being conducted in order to help anticipate issues regarding the comparability of the theoretical framework, constructs and the measurement of these constructs across groups. When research involves participants from distinct language groups, it is recommended that at least one, and preferably more, team members are fluent in the source language and the target language in which the survey will be administered.

Whenever possible, it is recommended to use measures that have been appropriately validated for the populations in which the questionnaire will be administered. Even when a measure has been validated within one population group, its validity may not extend to other groups, and additional steps may be necessary to increase validity and improve the value of comparisons across groups.

Translation of questionnaire items from one language to another should involve experienced translators. Review and adjudication of multiple, independent translations of the same items is currently considered the gold standard. If only one person translates the questionnaire, translation review should involve a group of bilingual people who are knowledgeable about questionnaire design principles and key study concepts. Translation assessment should not merely consist of back-translation.

Researchers should carefully select and translate items with the goal of achieving equivalence of construct meaning across study populations. In some cases, literal translation of a questionnaire item across linguistic variants of the survey will not adequately capture the construct of interest, and more flexible translation and adaptation of the question will be necessary.

All surveys, not just those that are translated, should be pre-tested to assess comprehension issues among the populations in which the survey will be administered. Ideally, pre-testing would involve cognitive interviewing before a survey is fielded. Cognitive interviewing or other pre-testing methods may also be used post-hoc to increase the validity of comparisons or to determine whether inconsistent results may be due to differential question comprehension.

Researchers should consider and seek solutions to minimise the ways in which culturally moderated response factors (e.g. social desirability, acquiescence, extreme responding) may influence responses.

Researchers should document decisions related to measurement development and item wording, especially where conceptual equivalence is suspect, translation is difficult, or where cognitive interviewing or other pre-testing methods reveals systematic differences in meaning. Researchers should also document issues around survey administration.

Outcomes and major determinants

Next, the Handbook presents constructs that are likely to be used across a range of policy evaluations, factors that can influence the validity of self-report tobacco use behaviours, factors that can influence comparability across surveys, and measures to assess use, providing examples from cross-national surveillance and evaluation systems as well as national sources. A core set of general mediator and moderator variables that may be relevant to consider in evaluations of tobacco control programmes and policies, with a brief description and assessment of some standard measures for assessing these constructs, are discussed. Self-report measures of nicotine/tobacco dependence in adults, concentrating on measures that are potentially appropriate for population-based/epidemiologic research, are reviewed as well.

Measuring Tobacco Use Behaviours

The Handbook describes the key concepts within the natural history of tobacco use, providing a conceptual model to guide measurement of key constructs. Current tobacco use is the most important construct because of its importance as an outcome in policy evaluation studies. Studies that have examined the validity of self-report measures of current use generally find these measures to be valid, although there exist some conditions under which the validity may be reduced.

It is important to measure the type of tobacco used, particularly in those countries in which a variety of types exist. The variety of forms available, the possibility of switching, or multiple concurrent use may influence the probability of quitting and disease risk.

Detailed measurement of information about tobacco product packaging is important in order to determine the variant of product type used, movement between price sectors and, potentially, to assess the use of tobacco from illicit sources.

Other important constructs in the measurement of tobacco use behaviour include early use, frequency and quantity of current use, quit attempts and duration of abstinence among former smokers.

Consumers of survey data in which tobacco use measures are included should be aware of factors that can influence population estimates of tobacco use, and take those into consideration when comparing estimates from surveys conducted within and across countries.

Measuring the Psychosocial Determinants of Tobacco Use and Dependence

The WG describes mediators and moderators theorised to be important in understanding how policies and interventions affect tobacco use behaviours, and under what circumstances they have an impact. A core set of measures likely to be important has been identified. Researchers should select from this list and, when appropriate, supplement it with other relevant measures, depending on the specific context and aims of each study. There are validated measures of many of the reviewed constructs, and researchers should whenever possible use these measures rather than developing their own ad hoc measures. Investigators should report the psychometric properties of their measurement instruments, reporting at least test-retest reliability, convergent validity and/or predictive validity. Psychological measures are particularly sensitive to wording and to cultural context, so we recommend that the methods for translations and cultural adaptations described elsewhere in the Handbook be utilised in populations where these measures have not been previously validated.

Measurement of nicotine dependence

Nicotine dependence is an important construct to assess as a moderator for the effects of tobacco control programmes and policies. The WG reviewed the evidence on the validity of various proposed measures of cigarette and smokeless-tobacco-induced nicotine dependence. For cigarette smoking, the 2-item Heaviness of Smoking Index is recommended for use in population-level studies. If only a single item measure is possible we would recommend the use of “time to first cigarette in the morning” as the item. For smokeless tobacco, the Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco (FTND-ST) appears to be a useful measure of nicotine dependence.

Existing data sources

The Handbook then describes sources of details about tobacco control policies, sources of information about tobacco production and trade and repo sitories of youth and adult surveillance surveys. These sources of information are particularly important for making comparisons between countries, and in some cases can be used to demonstrate policy impacts, although not the mechanisms by which they occur.

Data sources for monitoring tobacco control policies

The Handbook describes the new WHO Global Tobacco Control Report (GTCR), a repository of good-quality information on a wide range of tobacco control policies for the large majority of countries. The

GTCR contains copies of most of the legislation and regulations, some measures of scope and/or level of policy enactment, and an indicator of cases where national level policies may mask a diversity of sub-national policies. It is designed to be updated annually.

All policy researchers studying policy differences between countries should use it, and indeed it may be the easiest way to get this information for some individual countries.

The GTCR is limited in what it can provide on extent of implementation and/or enforcement. Its main limitation is that it does not contain information about sub-national policies, as information of this sort is only available for the limited number of countries that collect it.

Data sources on tobacco production, trade and sales

National data on the production, trade (export and import) and sales of tobacco products are most often available publicly at little to no cost and have been underutilised in evaluations of tobacco control programmes and policies. These data 1) can provide important insights into the relevant players and sectors in the national and regional political economy of tobacco control, 2) can be used to construct measures of historical trends in tobacco consumption and 3) provide estimates of the magnitude of the smuggling market. Thus, these data are important information sources for evaluation of tobacco control policies.

National data are typically available from sources such as government statistics agencies and ministries of trade and industry. The United Nations Statistical Division (UNSD) consolidates this information based on reports from countries. These reports are generally accurate, but primary sources should be used to confirm the data and to obtain other information such as data on sales and other tobacco products.

Data sources for monitoring global trends in tobacco use behaviours

The youth surveillance systems described in this section include The European School Survey Project on Alcohol and Other Drugs (ESPAD), the Global School-Based Student Health Survey (GSHS), the Global Youth Tobacco Survey (GYTS) and the Health Behaviour in School-Aged Children Survey (HBSC). The adult surveillance systems described include the Global Adult Tobacco Survey (GATS), the International Tobacco Control Policy Evaluation Survey (ITC) and the STEPwise Approach to Chronic Disease Factor Surveillance (STEPS).

To evaluate articles of the WHO FCTC among youth, GYTS is the only source of international data available that includes the following indicators: exposure to secondhand smoke, exposure to pro- and anti-tobacco media and advertising, cessation, minors' access and school curriculum.

To evaluate articles of the WHO FCTC among adults, GATS and ITC have the most comprehensive set of indicators, including: exposure to secondhand smoke, economics (price and taxation), cessation, product labelling, and exposure to pro- and anti-tobacco media and advertising. Where possible, longitudinal studies such as ITC should be used for evaluating policies and programmes because of the opportunity to examine and adjust for individual level predictors of tobacco use behaviours.

GYTS was developed, and GATS is being developed, for countries that did not have existing surveillance systems for the collection of information on tobacco use and its determinants.

Strategies for evaluating specific policy domains

The final section of the Handbook covers all major domains of tobacco control policies except for prevention policies and illicit trade. Here it is illustrated ways in which logic models can be used to highlight the different foci of policies. In particular, analysis of policy areas directed at controlling tobacco marketing (including some forms of product regulation) have identified the importance of monitoring tobacco industry innovations designed to mitigate the policy effects, while those less targeted at the industry have not done so.

Measures to Assess the Effectiveness of Tobacco Taxation

Article 6 of the WHO FCTC calls for ratifying nations to reduce the demand for tobacco products through taxation policies and other product price-related policies. This section focused on the measures needed for evaluating the impact of tobacco taxation, a highly effective tool for reducing tobacco use. The impact of tobacco taxes on tobacco use behaviours is mediated by tobacco product prices, tobacco company price-related marketing efforts, tobacco users' purchase behaviour, tax avoidance and smuggling.

Measuring tobacco product taxes is straightforward, with information on the level and structure of these taxes readily available from the Ministry of Finance and other sources (e.g. the International Monetary Fund, the WHO's GTCR). In some countries, it will also be important to measure sub-national taxes. Three methods for measuring tobacco product prices are discussed in this section: technology-based, observational and survey-based. These methods have differing strengths and weaknesses, and their costs will vary considerably.

To the extent that a national measure of price is of the most interest and a regularly repeated population survey of tobacco use is in place, including questions on price in such a survey would be most efficient. Measuring tobacco product purchase behaviour can be easily done through the addition of a limited set of questions to this survey. Developing accurate measures of tax avoidance and tobacco product smuggling is more challenging, and the validity of these measures is unclear and needs further research. Some of the questions on purchase behaviour in population surveys can be used to provide a range for the extent of tax avoidance. Multiple methods, most of which have not been widely applied and which need further research, can be used to assess the extent of tobacco product smuggling.

Measures to assess the effectiveness of smoke-free policies

Article 8 of the WHO FCTC calls for ratifying nations to adopt smoke-free policies for public indoor locations and workplaces. Evaluating the effects of public smoke-free policies is critical to understanding how these polices are implemented, whether they reduce exposure to tobacco smoke, and how they can be improved. The core constructs identified for evaluating smoke-free policies include compliance with the policy and exposure to tobacco smoke. Based on our review of the available research literature, we conclude that population surveys can generally be relied upon to provide valid measures of compliance with a public smoke-free policy and exposure to tobacco smoke. These self-report measures have been validated by ambient air monitoring and biomarkers of exposure to tobacco smoke. The review here also suggests that it may be important for evaluators to consider measuring key incidental effects of public smoke-free policies such as the impact on the behaviour of smokers, possible changes in smoking behaviour in the home and a variety of potential economic effects.

Measures to assess the effectiveness of tobacco product regulation

Articles 9 and 10 of the WHO FCTC call for ratifying nations to adopt policies for the regulation and disclosure of tobacco product contents and emissions. This section focuses on a review of the methods and measures for evaluating policies that are intended to regulate tobacco products. There are currently five main types:

1) regulations that require disclosure of product information;
2) regulations intended to reduce product toxicity and harm;
3) regulations intended to reduce the addictiveness and/or attractiveness of tobacco products;
4) regulations intended to prevent cigarette-caused fires; and
5) bans (or removal of bans) on product categories.

The selection of specific constructs and methods for evaluation will vary depending on the goals of the specific policy. However, as a general framework the impact of tobacco product

regulations on intended health outcomes will likely be moderated by changes in product design and performance, product marketing, product-related beliefs and attitudes, and tobacco use behaviour, which in turn are expected to influence exposures to tobacco constituents and emissions. Thus, evaluations should not be limited to assessing compliance within the intended effects of a regulation, but should also consider unintended effects or responses, such as tobacco industry innovation, that may interfere with the impact of the regulation.

There is a need for a centralised database that would at a minimum characterise different product regulations so that the effects of different policies can be compared. Additionally, as a condition permitting tobacco product sales, governments should require (if they do not currently do so) tobacco product manufacturers to regularly disclose information about their products at the finest level of brand subcategory, including sales and marketing data, product content and design features. This is needed to inform the development, implementation and evaluation of effective regulations. Additionally, ongoing surveillance is required to assess the impact of tobacco product regulation on the tobacco product market and on the population, as well as to detect industry responses and other unanticipated consequences of regulation. The challenges of measurement associated with evaluating the effects of tobacco product regulations should not be underestimated. For example, many governments have enacted maximum smoke emissions standards (i.e. tar, nicotine and carbon monoxide) based on standardised machine testing protocols for the purpose of reducing exposure to the constituents in tobacco products and resultant harm. However, based on the evidence reviewed in this Handbook, we recommend against using yields from standard machine testing protocols such as the ISO cigarette testing method (ISO Standard 3308, 2000) to assess or predict human exposure. Emission yields derived from these protocols are not valid measures of actual human exposure. In order to evaluate the effectiveness of product regulations aimed at reducing harm, measures of human use and exposure are essential. There is an urgent need to identify valid methods and measures for assessing human exposure and harm that have practical utility for evaluating tobacco product regulations.

Measures to assess the effectiveness of restrictions on tobacco marketing communications

Article 13 of the WHO FCTC encourages ratifying nations to adopt comprehensive tobacco marketing restrictions to the extent constitutionally possible. This section identifies the key issues and constructs for evaluating restrictions on tobacco marketing. Tobacco marketing includes all the communication efforts tobacco corporations use to encourage consumption of their products, including mass media advertising, sponsorship of sporting and cultural events, point of sale promotion, merchandising and give-aways, and public relations.

A core distinction to consider is between evaluation of the pathway of intended effects, and the need to monitor, and evaluate where necessary, evidence of tobacco industry activity that might reduce the impact of the policy.

Various methods can be used to measure the effects and effectiveness of restrictions on tobacco marketing, some borrowed from strategies to assess the impact of marketing. The main approaches include using consumer surveys to examine the target market's response to bans and restrictions and, if it can be obtained, use of disaggregated tobacco company marketing expenditure data to model changes in tobacco use. Given different limitations, the WG recommends a mix of these approaches, along with others where possible. However, there is a critical need to develop methods and valid measures for estimating the effects of marketing bans and restrictions at the level of the consumer.

Additional key challenges in evaluating the effects of marketing bans and restrictions include the extended time required for past marketing campaigns to dissipate from people's awareness, and the persistence of effects from recent campaigns. Innovative and increasingly subtle tobacco industry

marketing strategies create an urgent need for ongoing monitoring of industry behaviour.

Measures to assess effectiveness of product labelling

The WHO FCTC proposes tobacco product labelling regulations in 3 main areas: 1) health warnings, 2) misleading brand descriptors, such as “light” and “mild”, and 3) information on the constituents and emissions of tobacco products (Article 11). The Handbook identifies core constructs for evaluating labelling policies including: proximal outcomes such as awareness, processing and knowledge of health warnings; intermediate outcomes such as health knowledge, perceived risk, affective reactions, avoidance, brand appeal and cessation knowledge; and distal outcomes such as motivation to quit, changes in consumption patterns and quitting behaviours. Few of the measures for each of these constructs have undergone formal validation testing, although several of the measures described have shown utility for evaluating the impact of changes in product labelling.

The selection of specific measures to evaluate tobacco labelling policies will depend upon the policy chosen for evaluation. Evaluations of health warning labels should include proximal measures of noticing, along with intermediate measures of perceived risk or health knowledge. Evaluations of brand descriptors and other packaging elements should be a priority for tobacco control research. Unlike health warnings, these policies require the removal of information from the package and present challenges in the wording of survey measures. Evaluation of policies intended to communicate emissions and content information via packages should focus upon understanding and use of this information rather than knowledge or awareness.

Measuring the impact of anti-tobacco public communication campaigns

The WHO FCTC Article 12 requires ratifying countries to “promote and strengthen public awareness of tobacco control issues, using all available communication tools, as appropriate.” Such campaigns seek to increase awareness and knowledge of tobacco-related issues, with the goal of promoting individual behaviour change and support for and progress toward policy and social change. The Handbook provides a framework for evaluating multi-component public communication campaigns in order to design effective campaigns, identify and correct problems of campaigns that are in progress, and to document the campaign’s impact. Core methods include testing campaign messages during the design phase, monitoring the reach of the campaign during implementation, and assessing core constructs, including awareness, knowledge, attitudes and beliefs, support for policies and tobacco-related behaviour change. The measures described here, like the campaigns themselves, need to be customised to the specific content, purpose and message of the communication effort being implemented.

Regardless of the results of the public communication campaign (and particularly if it failed to show results), evaluations should be made publicly available. A system to collect and document campaign results would enhance our understanding both of how public communication campaigns work and how to make them better.

Measures to assess the effectiveness of tobacco cessation interventions

Article 14 of the WHO FCTC obligates ratifying nations to adopt policies that promote access to evidence-based tobacco cessation interventions. Such interventions range from less intensive efforts such as brief opportunistic advice by health care professionals to more intensive efforts delivered to tobacco users either individually or in groups by trained health professionals. Core constructs for evaluating access to tobacco cessation interventions include: proximal variables such as awareness of cessation interventions, intermediate variables including specific beliefs and attitudes about different cessation interventions, and distal variables reflecting the utilisation of different cessation interventions.

The effects of policies facilitating access to tobacco cessation interventions can be assessed through self-report using standardised surveys of current and former tobacco users and also by review of records that document trends in the utilisation of tobacco cessation interventions (e.g. calls to a helpline, sales of stop-smoking medications). Measures described here are useful exemplars of how to assess utilisation of cessation services. Evaluations of the effects of policies to promote access to cessation interventions should preferably employ a longitudinal design to assess the relationship between the utilisation of cessation treatments by current and former tobacco users and tobacco use behaviours.

Recommendations

Evaluation requires specific, committed resources. The framework the WG has developed highlights the potential value of good evaluation for interventions, as it allows for both ongoing improvement and the capacity to build on the accumulated knowledge acquired by others.

In 1999, the United States Centers for Disease Control and Prevention (CDC) recommended that 10% of the total budget for a comprehensive tobacco control programme should be allocated for evaluation and surveillance. The CDC recommendation was recently endorsed by WHO and represents a reasonable benchmark for governments to adopt.

The WG strongly recommends that countries allocate adequate funds for evaluation and surveillance activities. Where a budget for tobacco control programmes exists, we recommend that an adequate percentage of it be earmarked for evaluation and surveillance.

Evaluation needs to begin with an understanding of the nature of the interventions being evaluated. Collection of this information, especially for international studies, is surprisingly difficult. Collective effort, especially by agencies with networks into appropriate government agencies, could make it easier to collect this information, and do so in a consistent manner.

The WG recommends that high priority be given to the development and maintenance of a reliable and accurate international system for tracking tobacco control policies.

Also critical for the field to move forward is for sufficient attention and resources to be provided to knowledge utilisation, which in this domain would include appropriate detailed documentation of the results and all the features of evaluation studies, so as to allow the information to be compared and summative evaluations made. Development of a repository to collect and organise this information is becoming increasingly important. Complementing the repository of evaluations should be a similar repository of measures, with information as to their validity in the various contexts where they might be useful. The utility of such a repository would be enhanced by the development and agreement on use of prototype proformas for reporting on the validity data on measures, and on frequently repeated interventions, such as mass media campaigns. This will facilitate their combination into meta-analytic studies, especially important for understanding where and when things work. The continued momentum of the WHO FCTC and of the broader movement to fight against the global tobacco epidemic can be facilitated by the existence of such a repository, with appropriate tools for easy access and utilisation of the contents of the repository. Articles 20 and 22 of the WHO FCTC effectively call for such an initiative. Those conducting or sponsoring evaluations should be encouraged to add appropriate information to this repository.

The WG recommends that a repository be created and maintained to collect detailed documentation of the methods and results of tobacco control policy surveillance and evaluation studies, particularly those related to WHO FCTC policies.

The WG recommends that governments work together to support efforts to develop common methods and measures to support evaluations of tobacco control policies.

Governments should be encouraged to collect data from the tobacco industry to help evaluate current and future tobacco control policies, and to assist in identifying tobacco industry actions that might counteract the effects of

tobacco control policies. The kind of information that should be readily available from the industry and placed into the public repository includes disaggregated sub-brand specific marketing activities, product sales data and product content, design and performance data. It might also include more general information on political contributions, funding of scientists, general sponsorships and other activities of the industry that are designed to affect the environment in which they operate.

The WG recommends that governments mandate that tobacco companies provide them with information that might facilitate the evaluation of tobacco control policies and help identify the potential for new policies.

There are substantial infrastructure and information needs that are essential to conducting successful policy evaluations and supporting the dissemination and utilisation of evaluation results. Ongoing surveillance is required to assess the impact of tobacco control policies on the tobacco product market and on the population, as well as to detect industry responses to policies and other unanticipated consequences.

The WG recommends that countries interested in developing a tobacco control surveillance system be encouraged to join one of the international systems. Those countries that have existing national surveys are encouraged to link to these international efforts.

The information resources called for here should make important sources of data accessible and useable for informing policy, development, implementation and evaluation. Additionally, specific dissemination strategies should be employed to make relevant information useful to policy-makers, public health practitioners and the general public.

References

Abrams SM, Mahoney MC, Hyland A, et al. (2006). Early evidence on the effectiveness of clean indoor air legislation in New York State. *Am J Public Health*, 96(2):296-298.

Adam T, Mitschke S, Streibel T, et al. (2006). Quantitative puff-by-puff resolved characterization of selected toxic compounds in cigarette mainstream smoke. *Chem Res Toxicol*, 19(4):511-520.

Aftab M, Kolben D, Lurie P (1999). International cigarette labelling practices. *Tob Control*, 8(4):368-372.

Aguinis H, Pierce CA, Quigley BM (1993). Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: A meta-analytic review. *J Appl Soc Psychol*, 23(5):352-373.

Aitken PP, Leather DS, O'Hagan FJ (1985). Children's perceptions of advertisements for cigarettes. *Soc Sci Med*, 21:785-797.

Aitken PP, Leather DS, Squair SI (1986). Children's awareness of cigarette brand sponsorship of sports and games in the UK. *Health Educ Res*, 1(3):203-211.

Aitken PP, Eadie DR, Hastings GB, et al. (1991). Predisposing effects of cigarette advertising on children's intentions to smoke when older. *Br J Addict*, 86(4):383-390.

Ajzen I (1991). The theory of planned behavior. *Organ Behav Hum Decis Process*, 50(2):179-211.

Al-Delaimy WK (2002). Hair as a biomarker for exposure to tobacco smoke. *Tob Control*, 11(3):176-182.

Al-Delaimy WK, Mahoney GN, Speizer FE, et al. (2002). Toenail nicotine levels as a biomarker of tobacco smoke exposure. *Cancer Epidemiol Biomarkers Prev*, 11(11):1400-1404.

Albers AB, Siegel M, Cheng DM, et al. (2004). Relation between local restaurant smoking regulations and attitudes towards the prevalence and social acceptability of smoking: a study of youths and adults who eat out predominantly at restaurants in their town. *Tob Control*, 13(4):347-355.

Alciati MH, Frosh M, Green SB, et al. (1998). State laws on youth access to tobacco in the United States: measuring their extensiveness with a new rating system. *Tob Control*, (4):345-352.

Alechnowicz K, Chapman S (2004). The Philippine tobacco industry: "the strongest tobacco lobby in Asia." *Tob Control*, 13(Suppl 2):ii71-ii78.

Allwright S, Paul G, Greiner B, et al. (2005). Legislation for smoke-free workplaces and health of bar workers in Ireland: before and after study. *Br Med J*, 331(7525):1117.

Alterman AI, Gariti P, Cook TG, et al. (1999). Nicodermal patch adherence and its correlates. *Drug Alcohol Depend*, 53(2):159-165.

Altman DG, Levine DW, Coeytaux R, et al. (1996). Tobacco promotion and susceptibility to tobacco use among adolescents aged 12 through 17 years in a nationally representative sample. *Am J Public Health*, 86(11):1590-1593.

American Psychiatric Association (1995). *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, D.C., American Psychiatric Association.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (Text Revision)*. Washington, D.C., American Psychiatric Association.

Americans for Non-smokers Rights Foundation (2007). *Smoke-free lists, maps, and data*. (http://www.no-smoke.org).

Anderman C, Cheadle A, Curry S, et al. (1995). Selection bias related to parental consent in school-based survey research. *Eval Rev*, 19(6):663-674.

Anderson B, Silver B, Abramson P (1988). The effect of race of the interviewer on race-related attitudes of Black respondents in SRC/CPS national election studies. *Public Opin Q*, 52:289-324.

Anderson JR (1990). *Cognitive Psychology and its Implications*. New York, NY, WH Freedman.

Anderson KE, Kliris J, Murphy L, et al. (2003). Metabolites of a tobacco-specific lung carcinogen in nonsmoking casino patrons. *Cancer Epidemiol Biomarkers Prev*, 12(12):1544-1546.

Anderson P (2006). *A tool to assess the available services for smoking cessation at the country or region level*. Health Professionals and Smoking Cessation Project, 1-23. (*http://www.ensp.org/files/ 06_assessment_tool_for_smoking_cessation_hps2.pdf*)

Anderson P, Jané-Llopis E (2004). How can we increase the involvement of primary health care in the treatment of tobacco dependence? A meta-analysis. *Addiction*, 99(3):299-312.

Angst J, Gamma A, Neuenschwander M, et al. (2005). Prevalence of mental disorders in the Zurich Cohort Study: a twenty year prospective study. *Epidemiol Psichiatr Soc*, 14(2):68-76.

Ashley DL, Beeson MD, Johnson DR, et al. (2003). Tobacco-specific nitrosamines in tobacco from U.S. brand and non-U.S. brand cigarettes. *Nicotine Tob Res*, 5(3):323-331.

Ashley MJ, Cohen J, Ferrence R (2001). 'Light' and 'mild' cigarettes: who smokes them? Are they being misled? *Can J Public Health*, 92(6):407-411.

Assunta M, Chapman S (2004a). Industry sponsored youth smoking prevention pro-

gramme in Malaysia: a case study in duplicity. *Tob Control*, 13(Suppl 2):ii37-ii42.

Assunta M, Chapman S (2004b). "The world's most hostile environment": how the tobacco industry circumvented Singapore's advertising ban. *Tob Control*, 13(Suppl 2):ii51-ii57.

Atkin CK (2001). Theory and Principles of Media Health Campaigns. In: Rice RE, Foote DR, eds., *A Systems Based Evaluation Planning Model for Health Communication Campaigns in Developing Countries.* Thousand Oaks, CA, Sage Publications, 49-68.

Atkin CK, Freimuth VS (2001). Formative Evaluation Research in Campaign Design. In: Rice RE, Atkin CK, eds., *Public Communication Campaigns.* Thousand Oaks, CA, Sage Publications, 125-145.

Babor TF, Brown J, Del Boca FK (1990). Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Behav Assess*, 12(1):5-31.

Baker TB, Conti DV, Moffit T, et al. (*in press*). The nicotine dependence phenotype: Translating theoretical perspectives and extant data into recommendations for measurement. In: *Phenotypes, Endophenotypes, and Genetic Studies of Nicotine Dependence*. Smoking and Tobacco Control Monograph No. 22, Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute.

Bakkalbasi N, Bauer K, Glover J, et al. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed Digit Libr*, 3:7.

Balch G, Sutton SM (1997). Keep Me Posted. A Plea for Practical evaluation. In: Goldberg ME, Fishbein M, Middlestadt SE, eds., *Social Marketing: Theoretical and Practical Perspectives.* Mahwah, Lawrence Erlbaum Associates, 61-74.

Bandura A (1986). *Social Foundations of Thought and Action: A Social-Cognitive Theory.* Englewood Cliffs, Prentice-Hall.

Barnard M, Forsyth A (1996). The social context of under-age smoking: A qualitative study of cigarette brand preference. *Health Educ J*, 55(2):175-184.

Barnes DE, Bero LA (1997). Scientific quality of original research articles on environmental tobacco smoke. *Tob Control*, 6(1):19-26.

Barnes DE, Bero LA (1998). Why review articles on the health effects of passive smoking reach different conclusions. *JAMA*, 279(19):1566-1570.

Baron RM, Kenny DA (1986). The moderator mediator variable distinction in social psychological research - conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*, 51(6):1173-1182.

Barrett PT, Petrides KV, Eysenck SGB, et al. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries**.** *Pers Individ Dif*, 25(5):808-819.

Barton J (1998). *Young Teenagers and Smoking in 1997: A Report of the Key Findings from the Teenage Smoking Attitudes Survey Carried out in England*. London, Health Education Authority.

Bates B, Blenkinsop S, Boreham R, et al. (2005). *Smoking, drinking and drug use among young people in England 2004*. Leeds, Health and Social Care Information Centre.

Bates C (1999). Tobacco sponsorship of sports. *Br J Sports Med*, 33(5):299-300.

Bauer JE, Hyland A, Li Q, et al. (2005). A longitudinal assessment of the impact of smoke-free worksite policies on tobacco use. *Am J Public Health*, 95(6):1024-1029.

Bauer UE, Johnson TM (1999). *Assessing program impacts, 1998 to 1999. Florida Youth Tobacco Survey (FYTS).* Tallahassee, Florida Department of Health, Bureau of Epidemilogy. Vol. 2, Report 4:20. *(http://edocs. dlis.state.fl.us/fldocs/doh/tobacco/fyts/vol2rep_4.pdf)*

Bauer UE, Johnson TM (2000). Editing data: what difference do consistency checks make? *Am J Epidemiol*, 151(9): 921-926.

Bauld L, Coleman T, Adams C, et al. (2005). Delivering the English smoking treatment services. *Addiction*, 100(Suppl 2):19-27.

Baumgartner H, Steenkamp J (2001). Response styles in marketing research: A cross-national investigation. *J Mark Res*, 38:143-156.

Beaton DE, Bombardier C, Guillemin F, et al. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24): 3186-3191.

Becker MH (1974). *The Health Belief Model and Personal Health Behavior*. Thorofare, NJ, Charles B. Stock.

Beede P, Lawson R (1992). The effect of plain packages on the perception of cigarette health warnings. *Public Health*, 106(4):315-322.

Beltramini RF (1988). Perceived believability of warning label information presented in cigarette advertising. *J Advert*, 17:26-32.Benowitz NL (1996a). Cotinine as a biomarker of environmental tobacco smoke exposure. *Epidemiol Rev*, 18(2): 188-204.

Benowitz NL (1996b). Biomarkers of cigarette smoking. In: Shopland DR, ed. *The FTC Cigarette Test Method for Determining Tar, Nicotine, and Carbon Monoxide Yields of U.S. Cigarettes*. Report of the NCI Expert Committee. Smoking and Tobacco Control Monograph No. 7, Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub No. 96-4028; 93-111.

Benowitz NL (1998). *Nicotine Safety and Toxicity.* New York,NY. Oxford University Press.

Benowitz NL (1999). Biomarkers of environmental tobacco smoke exposure. *Environ Health Perspect*, 107(Suppl 2):349-355.

Benowitz NL (2003). Cigarette smoking and cardiovascular disease: pathophysiology and implications for treatment. *Prog Cardiovasc Dis*, 46(1):91-111.

Benowitz NL, Henningfield JE (1994). Establishing a nicotine threshold for addiction. The implications for tobacco regulation. *N Engl J Med*, 331(2):123-125.

Benowitz N, Hall S, Herning R, et al. (1983). Smokers of low-yield cigarettes do not consume less nicotine. *N Engl J Med*, 309(3):139-142.

Benowitz NL, Perez-Stable EJ, Herrera B, et al. (2002). Slower metabolism and reduced intake of nicotine from cigarette smoking in Chinese-Americans. *J Natl Cancer Inst*, 94(2):108-115.

Berkowitz B (2001). Studying the outcomes of community-based coalitions. *Am J Commun Psychol*, 29(2):213-227.

Bernard HR (1994). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Walnut Creek, Altmira.

Bero LA (2005). Tobacco industry manipulation of research. *Public Health Rep*, 120(2):200-208.

Bialous SA, Yach D (2001). Whose standard is it, anyway? How the tobacco industry determines the International Organization for Standardization (ISO) standards for tobacco and tobacco products. *Tob Control*, 10(2):96-104.

Biener L, Harris JE, Hamilton W (2000). Impact of the Massachusetts tobacco control programme: population based trend analysis. *Br Med J*, 321(7257):351-354.

Biener L, Garrett CA, Gilpin EA, et al. (2004). Consequences of declining survey response rates for smoking prevalence estimates. *Am J Prev Med*, 27(3):254-257.

Bierut LJ, Madden PA, Breslau N, et al. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*, 16(1):24-35.

Billiet J, McClendon M (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Struct Equation Model*, 7:608-628.

Blackford AL, Yang G, Hernandez-Avila M, et al. (2006). Cotinine concentration in smokers from different countries: relationship with amount smoked and cigarette type. *Cancer Epidemiol Biomarkers Prev*, 15(10):1799-1804.

Blecher EH (2006). The effects of the tobacco products control amendment act of 1999 on restaurant revenues in South Africa: a panel data approach. *S Afr J Econ*, 74(1).

Blecher EH, van Walbeek CP (2004). An international analysis of cigarette affordability. *Tob Control*, 13(4):339-346.

Blumberg SJ, Luke JV, Cynamon ML (2006). Telephone coverage and health survey estimates: evaluating the need for concern about wireless substitution. *Am J Public Health*, 96(5):926-931.

Bollen KA (1989). *Structural Equations with Latent Variables*. New York, NY, John Wiley & Sons.

Bollen KA (2002). Latent variables in psychology and the social sciences. *Annu Rev Psychol*, 53:605-634.

Bollen KA, Entwisle B, Alderson AS (1993). Macrocomparative research methods. *Annu Rev Sociol*, 19:321-351.

Bonsignore M, Barkow K, Jessen F, et al. (2001). Validity of the five-item WHO Well-Being Index (WHO-5) in an elderly population. *Eur Arch Psychiatry Clin Neurosci*, 251(Suppl 2):II27-II31.

Boreham R, Shaw A (2001). *Smoking, Drinking and Drug Use Among Young People in Scotland in 2000*. Edinburgh, The Stationery Office. (*http://www.scotland.gov.uk/Publications/2001/11/10381/File-1*)

Borgerding JA, Bodnar JA, Wingate DE. (2000). *The 1999 Massachusetts Benchmark Study: Final Report*. Louisville, KY, Brown and Williamson Tobacco.

Borland R (2003). A strategy for controlling the marketing of tobacco products: a regulated market model. *Tob Control*, 12(4): 374-382.

Borland R, Hill D (1997a). Initial impact of the new Australian tobacco health warnings on knowledge and beliefs. *Tob Control*, 6(4):317-325.

Borland R, Hill D (1997b). The path to Australia's tobacco health warnings. *Addiction*, 92(9):1151-1157.

Borland R, Mullins R, Trotter L, et al. (1999). Trends in environmental tobacco smoke restrictions in the home in Victoria, Australia. *Tob Control*, 8(3):266-271.

Borland R, Yong HH, King B, et al. (2004). Use of and beliefs about light cigarettes in four countries: findings from the International Tobacco Control Policy Evaluation Survey. *Nicotine Tob Res*, 6(Suppl 3):S311-S321.

Borland R, Yong HH, Cummings KM, et al. (2006a). Determinants and consequences of smoke-free homes: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15(Suppl 3):iii42-iii50.

Borland R, Yong HH, Siahpush M, et al. (2006b). Support for and reported compliance with smoke-free restaurants and bars by smokers in four countries: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15(Suppl 3):iii34-iii41.

Bornstein RF (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychol Bull*, 106(2):265-289.

Borsch-Supan A, Brugiavini A, Jürges H, et al. (2005). *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe, (SHARE)*. Mannheim, Mannheim Research Institute for the Economics of Aging. *(http://www.share-project.org)*

Box GEP, Jenkins GM, Reinsel GC (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, Prentice-Hall.

Boyle RG, Jensen J, Hatsukami DK, et al. (1995). Measuring dependence in smokeless tobacco users. *Addict Behav*, 20(4):443-450.

Bracht N (2001). Community partnership strategies in health campaigns. In: Rice RE, Atkin CK, eds., *Public Communication Campaigns.* Thousand Oaks, CA, Sage Publications, 119-128.

Bradburn N, Sudman S, Wansink B (2004). *Asking Questions.* San Francisco, Jossey-Bass.

Bradford JA, Harlan WR, Hanmer HR (1936). Nature of cigarette smoke. Technic of experimental Smoking**.** *Ind Eng Chem*, 28(7):836-839.

Braverman MT, Aarø LE (2004). Adolescent smoking and exposure to tobacco marketing under a tobacco advertising ban: findings from 2 Norwegian national samples. *Am J Public Health*, 94(7):1230-1238.

BRC Marketing & Social Research (2004). *Smoking Health Warnings Study: The Effectiveness of Different (Pictorial) Health Warnings in Helping People Consider their Smoking-Related Behaviour.* Wellington, BRC Marketing & Social Research.

Breland AB, Kleykamp BA, Eissenberg T (2006). Clinical laboratory evaluation of potential reduced exposure products for smokers. *Nicotine Tob Res*, 8(6):727-738.

Brener ND, Collins JL, Kann L, et al. (1995). Reliability of the Youth Risk Behavior Survey Questionnaire. *Am J Epidemiol*, 141(6):575-580.

Brener ND, Billy JO, Grady WR (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *J Adolesc Health*, 33(6):436-457.

Brener ND, Grunbaum JA, Kann L, et al. (2004). Assessing health risk behaviors among adolescents: the effect of question wording and appeals for honesty. *J Adolesc Health*, 35(2):91-100.

Brener ND, Eaton DK, Kann L, et al. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opin Q*, 70(3):354-374.

Breslow NE, Day, NE (1980) *IARC Scientific Publications No. 32, Statistical Methods in Cancer Research*, Vol.2, *The Analysis of Cohort Studies,* Lyon, IARC

Breslow, NE, Day, NE (1987) *IARC Scientific Publications No. 82, Statistical methods in Cancer Research*, Vol. 2, *The Design and Analysis of Cohort Studies*, Lyon, IARC

Breslau N (1995). Psychiatric comorbidity of smoking and nicotine dependence. *Behav Genet*, 25(2):95-10.

Breslau N, Johnson EO (2000). Predicting smoking cessation and major depression in nicotine-dependent smokers. *Am J Public Health*, 90(7):1122-1127.

Breslau N, Novak SP, Kessler RC (2004). Psychiatric disorders and stages of smoking**.** *Biol Psychiatry*, 55(1):69-76.

Breteler MH, Hilberink SR, Zeeman G, et al. (2004). Compulsive smoking: the development of a Rasch homogeneous scale of nicotine dependence. *Addict Behav*, 29(1):199-205.

Brislin RW (1970). Back-translation for cross-cultural research. *J Cross Cult Psychol*, 1:186-216.

Brittingham A, Tourangeau R, Kay W. (1998). Reports of smoking in a national survey: data from screening and detailed interviews, and from self- and interviewer-administered questions. *Ann Epidemiol*, 8(6):393-401.

Broadstock M. (2007). Systematic review of the health effects of modified smokeless tobacco products. *New Zealand Health Technol Assess Rep,* 10(1):129.

Broms U, Madden PA, Heath AC, et al. (2007). The Nicotine Dependence Syndrome Scale in Finnish smokers. *Drug Alcohol Depend*, 89(1):42-51.

Brown KS, Diener A, Ahmed R, et al. (2005). Survey Methods. In: *2002 Youth Smoking Survey Technical Report*. Ottawa, Health Canada. *(http://www.hc-sc.gc.ca/hecs-sesc/ tobacco/research/yss-sjf2002/chap2_e. html).*

Brubaker RG, Mitby SK (1990). Health-risk warning labels on smokeless tobacco products: are they effective? *Addict Behav*, 15(2):115-118.

Brunnemann KD, Hoffmann D (1974). Chemical studies on tobacco smoke. XXV. The pH of tobacco smoke. *Food Cosmet Toxicol*, 12(1):115-124.

Brunnemann KD, Hoffman JM (1982). Pyrolytic origins of major gas phase constituents of gigarette smoke. *Recent Adv Tob Sci*, 8:103.

Brunnemann KD and D. Hoffmann. Pyrolytic origins of major gas phase constituents of cigarette smoke. *Recent Adv. Tob. Sci.* 8:103-140. 1982

Brunnemann KD, Kagan MR, Cox JE, et al. (1990). Analysis of 1,3-butadiene and other selected gas-phase components in cigarette mainstream and sidestream smoke by gas chromatography-mass selective detection. *Carcinogenesis*, 11(10):1863-1868.

Bryant MS, Skipper PL, Tannenbaum SR, et al. (1987). Hemoglobin adducts of 4-aminobiphenyl in smokers and nonsmokers. *Cancer Res*, 47(2):602-608.

Burns DM, Anderson CM, Major J, et al. (2000). Cessation and cessation measures among adult daily smokers: National and state-specific data. In: Shopland DR, Burns DM, Amacher RH and Ruppert W, eds. *Population Impact of Smoking Cessation: Proceedings of a Conference on What Works to Influence Cessation in the General Population.* Smoking and Tobacco Control Monograph No. 12. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub. No. 00-4892; 25-98.

Burns DM, Major JM, Shanks TG, et al. (2001). Smoking lower yield cigarettes and disease risks. In: Shopland DR, Burns DM,

Benowitz NL, and Amacher RH, eds. *Risks Associated with Smoking Cigarettes with Low Machine-Measured Yields of Tar and Nicotine.* Smoking and Tobacco Control Monograph No. 13, Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub No. 02-5074; 65-158.

Bush K, Kivlahan DR, McDonell MB, et al. (1998). The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med*, 158(16): 1789-1795.

Bush LP, Cui M, Shi H, et al. (2001). Formation of tobacco-specific nitrosamines in air-cured tobacco. *Recent Adv Tob Sci*, 27:23-46.

Byrne DG, Mazanov J (2001). Self-esteem, stress and cigarette smoking in adolescents**.** *Stress Health*, 17(2):105-110.

Byrne F, Chapman S (2005). The most cited authors and papers in tobacco control. *Tob Control*, 14(3):155-160.

Calafat AM, Polzin GM, Saylor J, et al. (2004). Determination of tar, nicotine, and carbon monoxide yields in the mainstream smoke of selected international cigarettes. *Tob Control*, 13(1):45-51.

Campbell IA, Prescott RJ, Tjeder-Burton SM (1996). Transdermal nicotine plus support in patients attending hospital with smoking-related diseases: a placebo-controlled study. *Respir Med*, 90(1):47-51.

Canadian Cancer Society (2001). *Evaluation of New Warnings on Cigarette Packages*. Toronto, Environics, Focus Canada.

Cannon DS, Baker TB, Piper ME, et al. (2005). Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. *Nicotine Tob Res*, 7(6):853-858.

Cannon T (1992). *Basic Marketing.Principles and Practice*. London, Cassell.

Capehart T (2007). Harvest intentions for 2007-crop tobacco advance 2 percent. *Tobacco Outlook,* TBS 262. United States Department of Agriculture.

Caraballo RS, Giovino GA, Pechacek TF, et al. (1998). Racial and ethnic differences in serum cotinine levels of cigarette smokers: Third National Health and Nutrition Examination Survey, 1988-1991. *JAMA*, 280(2):135-139.

Caraballo RS, Giovino GA, Pechacek TF, et al. (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years or older: Third National Health and Nutrition Examination Survey, 1988-1994. *Am J Epidemiol*, 153(8):807-814.

Caraballo RS, Giovino GA, Pechacek TF (2004). Self-reported cigarette smoking vs. serum cotinine among U.S. adolescents. *Nicotine Tob Res*, 6(1):19-25.

Carpenter CM, Wayne GF, Pauly JL, et al. (2005). New cigarette brands with flavors that appeal to youth: tobacco marketing strategies. *Health Aff (Millwood)*, 24(6): 1601-1610.

Carr-Gregg MRC, Gray AJ (1993). Generic packaging: A possible solution to the marketing tobacco to young people. *World Smoking and Health*, 18(2):11-13.

Cecil H, Evans RI, Stanley MA (1996). Perceived believability among adolescents of health warning labels on cigarette packs. *J Appl Soc Psychol*, 26:502-519.

Center for Media Education (1997). *Alcohol and Tobacco on the Web, New Threats to Youth*. Washington, DC, Center for Media Education.

Center for Media Education. (1998). Tobacco *Targeted at Youth on the Internet: An update.* Washington, DC, Center for Media Education.

Center for Media Education (2007). *Tobacco targeted at youth on the internet: an update*. Washington, DC, Center for Media Education.

Centers for Disease Control and Prevention (1991). Differences in the age of smoking initiation between blacks and whites—United States. *Morb Mortal Wkly Rep*, 40(44):754-757.

Centers for Disease Control and Prevention (1993). Smoking Cessation During Previous Year Among Adults - United States, 1990 and 1991. *Morb Mortal Wkly Rep*, 42 (26):504-507.

Centers for Disease Control and Prevention (1994a). Cigarette smoking among adults—United States, 1992, and changes in the definition of current cigarette smoking. *Morb Mortal Wkly Rep,* 43(19):342-346.

Centers for Disease Control and Prevention (1994b). Reasons for tobacco use and symptoms of nicotine withdrawal among adolescent and young adult tobacco users—United States, 1993. *Morb Mortal Wkly Rep*, 43(41):745-750.

Centers for Disease Control and Prevention (1994c). Changes in the cigarette brand preferences of adolescent smokers - United States, 1989-1993. *JAMA*, 272:843-844.

Centers for Disease Control and Prevention (1997a). Protocol to measure the quantity of nicotine contained in smokeless tobacco products manufactured, imported, or packaged in the United States. *Fed Regist*, 62(85):24115-24117.

Centers for Disease Control and Prevention (1997b). Filter ventilation levels in selected U.S. cigarettes, 1997. *Morb Mortal Wkly Rep*, 46(44):1043-1047.

Centers for Disease Control and Prevention (1998). Incidence of initiation of cigarette smoking—United States, 1965-1996. *Morb Mortal Wkly Rep*, 47(39):837-840.

Centers for Disease Control and Prevention (1999a). *Best Practices for Comprehensive Tobacco Control Programs-August 1999*. Atlanta, GA, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Centers for Disease Control and Prevention (1999b). *Handbook for Conduction Youth Risk Behavior Surveys*. Atlanta,GA, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Centers for Disease Control and Prevention (2001). *Global Youth Tobacco Survey (GYTS), 2001 Handbook*. Atlanta, GA, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Centers for Disease Control and Prevention (2003). *Designing and Implementing an Effective Tobacco Countermarketing Campaign*. Atlanta, GA, Centers for Disease Control Prevention, U.S. Department of Public Health and Human Services, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Centers for Disease Control and Prevention (2004). *Telephone Quitlines: A Resource for Development, Implementation, and Evaluation*. Atlanta, GA, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Centers for Disease Control and Prevention (2005). Annual smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 1997-2001. *Morb Mortal Wkly Rep*, 54(25):625-628.

Centers for Disease Control and Prevention (2007). *State Tobacco Activities Tracking and Evaluation (STATE) System*. (http://apps nccd cdc gov/statesystem)

Centre for Social Marketing (1995). *Smoking and Deprivation Study Part 2: An Investigation into Cigarette Coupon Schemes*. Glasgow, Centre for Social Marketing, University of Strathclyde.

Centre for Tobacco Control Research (2001). *Tobacco Brand Stretching: A Report prepared for the Department of Health*. Glasgow, Centre for Tobacco Control Research, University of Strathclyde.

Cepeda-Benito A, Henry K, Gleaves D, et al. (2004). Cross-cultural investigation of the questionnaire of smoking urges in American and Spanish smokers. *Assessment*, 11:152-159.

Chabrol H, Niezborala M, Chastan E, et al. (2003). A study of the psychometric properties of the Fagestrom Test for Nicotine Dependence. *Addict Behav*, 28(8):1441-1445.

Chaloupka F, Hu TW, Warner KE, et al. (2000a). The taxation of tobacco products. In: Jha P, Chaloupka FK, eds., *Tobacco Control in Developing Countries.* Oxford, Oxford University Press.

Chaloupka F, Tauras JA, Grossman M (2000b). The Economics of Addiction. In: Jha P, Chaloupka FK, eds., *Tobacco Control in Developing Countries.* Oxford, Oxford University Press.

Chaloupka FJ (*in press*). State and local policies. In: Haviland L. ed., *Lighting a Fuse for Public Health: Tobacco Control Lessons Learned*. Washington, DC, JSI/American Legacy Foundation.

Chaloupka FJ, Laixuthai A (1996). *U.S. Trade Policy and Cigarette Smoking in Asia*. Cambridge, MA, National Bureau of Economic Research, Inc.

Chaloupka FJ, Wakefield M, Czart C (2001). Taxing Tobacco: The Impact of Tobacco Taxes on Cigarette Smoking and Other Tobacco Use. In: Rabin RL, Sugarman SD, eds., *Regulating Tobacco.* Oxford, Oxford University Press.

Chaloupka FJ, Cook PJ, Peck RM, et al. (2008). *Enhancing compliance with tobacco control policies*. Sanford working papers series, Terry Sanford Institute of public policy. Durham, NC, Sanford working papers series, Terry Sanford Institute of public policy: SAN08-07.

Chapman S (1989). The limitations of econometric analysis in cigarette advertising studies. *Br J Addict*, 84(11):1267-1274.

Chapman S (1992). Changes in adult cigarette consumption per head in 128 countries, 1986-90. *Tob Control*, 1:281-284.

Chapman S, Davis RM (1997). Smoking in movies: is it a problem? *Tob Control*, 6(4):269-271.

Chapman S, Dominello A (2001). A strategy for increasing news media coverage of tobacco and health in Australia**.** *Health Promot Int*, 16(2):137-143.

Chapman S, Carter M (2003). Avoid health warnings on all tobacco products just as long as we can: a history of Australian tobacco industry efforts to avoid, delay and dilute health warnings on cigarettes. *Tob Control*, 12:13-22.

Chapman S, Wilson D, Wakefield M (1986). Smokers' understandings of cigarette yield labels. *Med J Aust*, 145(8):376-379.

Chapman S, Wong WL, Smith W (1993). Self-exempting beliefs about smoking and health: differences between smokers and ex-smokers. *Am J Public Health*, 83(2): 215-219.

Charlton A (1986). Children's advertisement awareness related to their views on smoking. *Health Educ J*, 45(2):75-78.Charlton A, While D, Kelly S (1997). Boys' smoking and cigarette-brand-sponsored motor racing. *Lancet*, 350:1474.

Chesterman J, Judge K, Bauld L, et al. (2005). How effective are the English smoking treatment services in reaching disadvantaged smokers? *Addiction*, 100(Suppl 2): 36-45.

Cheung G, Rensvold R (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *J Cross Cult Psychol*, 31:187-212.

Choi WS, Gilpin EA, Farkas AJ, et al. (2001). Determining the probability of fu-

ture smoking among adolescents. *Addiction*, 96(2):313-323.

Christie D, Etter JF (2004). Utilization and impact of cigarette pack covers illustrated with antismoking messages. *Eval Health Prof*, 27:107-118.

Christie DH, Etter JF (2005). Validation of English-language versions of three scales measuring attitudes towards smoking, smoking-related self-efficacy and the use of smoking cessation strategies. *Addict Behav*, 30(5):981-988.

Cialdini RB (2003). Crafting normative messages to protect the environment**.** *Curr Dir Psychol Sci*, 12(4):105-109.

Clark DB, Wood DS, Martin CS, et al. (2005). Multidimensional assessment of nicotine dependence in adolescents. *Drug Alcohol Depend*, 77(3):235-242.

Clark LA, Watson D (1995). Constructing validity: basic issues in objective scale development. *Psychol Assess*, 7:309-319.

Claude T (1973). *Research Planning Memorandum on Some Thoughts about New Brands of Cigarettes for the Youth Market*. FTC Joe Camel document. Bates No: 505101981-1992 *(http://tobaccodocuments. org/ ftc_rjr/CX000043.html).*

Clegg Smith K, McElrath YT, Wakefield M, et al. (2005). Media advocacy and newspaper coverage of tobacco issues: A comparative analysis of 1 year's print news in the United States and Australia. *Nicotine Tob Res*, 7(2):289-299.

Cloninger CR, Svrakic DM, Przybeck TR (1993). A psychobiological model of temperament and character. *Arch Gen Psychiatry*, 50(12):975-990.

Coeytaux RR, Altman DG, Slade J (1995). Tobacco promotions in the hands of youth**.** *Tob Control*, 4:253-257.

Coffman J (2002). *Public Communication Campaign Evaluation: An Environmental Scan of Challenges, Criticisms, Practice, and Opportunities*. Cambridge, Harvard Family Research Project.

Cogliano, VJ, Baan, RA, Straif, K. et al. (2004) The science and practice of carcinogen identification and evaluation. *Environmental Health Perspectives,* 112: 1269-1274.

Cohen JB (1996a). Smokers' knowledge and understanding of advertised tar numbers: health policy implications. *Am J Public Health*, 86(1):18-24.

Cohen JB (1996b). Consumer/smoker perceptions of Federal Trade Commission Tar Ratings. In: Shopland DR, ed. *The FTC Cigarette Test Method for Determining Tar, Nicotine, and Carbon Monoxide Yields of US Cigarettes*. Report of the NCI Expert Committee. Smoking and Tobacco Control Monograph No.7, Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub No. 96-4028:127-134.

Cohen S, Kamarck T, Mermelstein R (1983). A global measure of perceived stress. *J Health Soc Behav*, 24(4):385-396.

Cohen SJ, Katz BP, Drook CA, et al. (1988). Overreporting of smokeless tobacco use by adolescent males. *J Behav Med*, 11(4):383-393.

Coleman T, Cheater F, Murphy E (2004). Qualitative study investigating the process of giving anti-smoking advice in general practice. *Patient Educ Couns*, 52(2):159-163.

Coleman T, Pound E, Adams C, et al. (2005). Implementing a national treatment service for dependent smokers: initial challenges and solutions. *Addiction*, 100(Suppl 2):12-18.

Collett Dickenson Pearce (1996). *Benson & Hedges Special Filter: The Biggest Brand in the UK*. Document No. 0139:1-42.

Collett Dickenson Pearce (1998). *Berkley Thoughts & Ideas: Some Views from Soho Square*. Document 0280:1-5.

Collett Dickenson Pearce (1999). *Correspondence, May and June 1999*. Document No. 0346.

Commomwealth of Massachusetts DoPH. (2007). Cigarette and Smokeless Tobacco Products: Reports of Added Constituents and Nicotine Ratings (105 CMR 660.000). (*http://www mass gov/dph/mtcp/legal/ingreg htm*)

Connell JP, Kubisch AC (1998). Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems. In: Fulbright-Anderson K, Kubisch AS, Connell JP, eds., *New Approaches to Evaluating Community Initiatives, Volume 2: Theory, Measurement, and Analysis.* Washington, D.C., The Aspen Institute of Human Studies. *(http://www.aspeninstitute.org/ site/c.huLWJeMRKpH/b.613709/k. B547/ Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives Progress Prospects and Problems.htm).*

Connell JP, Kubish AC, Schorr LB, et al. (2007). *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington, D.C., The Aspen Institute of Human Studies.

Conner M, Norman P (1996). *Predicting Health Behaviour: Research and Practice with Social Cognition Models*. Buckingham, Open University Press.

Connolly GN (1995). The marketing of nicotine addiction by one oral snuff manufacturer. *Tob Control*, 4(1):73-79.

Connolly GN, Alpert HR, Rees V, et al. (2005). The Effect of the New York State Cigarette Fire Safety Standard on Ignition Propensity, Smoke Toxicity, and the Consumer Market. *Tob Control*, 14:321-324.

Connolly GN, Alpert HR, Wayne GF, et al. (2007). *Trends in Smoke Nicotine Yield and Relationship to Design Characteristics among Popular U.S. Cigarette Brands, 1997-2005*. A Report of the Tobacco Research Program Division of Public Health Practice. Boston, MA, Harvard University School of Public Health.

Conrad FG, Blair J (2004). Data Quality in Cognitive Interviews: The Case of Verbal Reports. In: Pressler S, Rothgeb JM,

Couper MP, et al., eds., *Methods for Testing and Evaluating Survey Questionnaires.* Hoboken, NJ, Wiley & Sons, 67-88.

Conrad KM, Flay BR, Hill D (1992). Why children start smoking cigarettes: predictors of onset. *Br J Addict*, 87(12):1711-1724.

Conway M, Ross M (1984). Getting what you want by revising what you had. *J Pers Soc Psychol*, 47(4):738-748.

Cook TD, Campbell DT (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA, Houghton Mifflin.

Cotter P, Cohen J, Coulter P (1982). Race-of-interviewer effects in telephone interviews**.** *Public Opin Q*, 46:278-284.

Cowling DW, Bond P (2005). Smoke-free laws and bar revenues in California—the last call. *Health Econ*, 14(12):1273-1281.

Cowling DW, Johnson TP, Holbrook BC, et al. (2003). Improving the self reporting of tobacco use: results of a factorial experiment. *Tob Control*, 12(2):178-183.

CREATEC and Market Studies (2003). Effectiveness of Health Warning Messages on Cigarette Packages in Informing Less-Literate Smokers. Montreal, Quebec, Communication Canada. *(http://www.smokefree.ca/ warnings/WarningsResearch/Createc_LiteracyFinal_EN__Dec03.doc)*

Critchley JA, Unal B (2003). Health effects associated with smokeless tobacco: a systematic review. *Thorax*, 58(5):435-443.

Cummings KM, Morley CP, Horan JK, et al. (2002a). Marketing to America's youth: evidence from corporate documents. *Tob Control*, 11(Suppl 1):I5-17.

Cummings KM, Morley CP, Hyland A (2002b). Failed promises of the cigarette industry and its effect on consumer misperceptions about the health risks of smoking. *Tob Control*, 11(Suppl 1):I110-I117.

Cummings KM, Hyland A, Perla J, et al. (2003). Is the prevalence of youth smoking affected by efforts to increase retailer compliance with a minors' access law? *Nicotine Tob Res*, 5(4):465-471.

Dale LC, Glover ED, Sachs DP, et al. (2001). Bupropion for smoking cessation: predictors of successful outcome. *Chest*, 119(5):1357-1364.

Dalton MA, Sargent JD, Beach ML, et al. (2003). Effect of viewing smoking in movies on adolescent smoking initiation: a cohort study. *Lancet*, 362(9380):281-285.

Davis D (1997). Nonrandom measurement error and race of interviewer effects among African Americans. *Public Opin Q*, 61:183-207.

Davis RM (1991). Reducing youth access to tobacco. *JAMA*, 266(22):3186-3188.

Davis RM, Healy P, Hawk SA (1990). Information on tar and nicotine yields on cigarette packages. *Am J Public Health*, 80(5):551-553.

de Leon J, Becoña E, Gurpegui M, et al. (2002). The association between high nicotine dependence and severe mental illness may be consistent across countries. *J Clin Psychiatry*, 63(9):812-816.

De Vries H, Dijkstra M, Kuhlman P (1988). Self-efficacy: the third factor besides attitude and subjective norm as a predictor of behavioural intentions. *Health Educ Res*, 3:273-282.

De Vries H, Backbier E, Kok G, et al. (1995). The impact of social influences in the context of attitude, self-efficacy, intention, and previous behavior as predictors of smoking onset. *J Appl Soc Psychol*, (25):237-257.

DeMarini DM (2004). Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutat Res*, 567 (2-3): 447-474.

Dent CW, Galaif J, Sussman S, et al. (1993). Demographic, psychosocial and behavioral differences in samples of actively and passively consented adolescents. *Addict Behav*, 18(1):51-56.

Department of Health (2003). Permitted additives to tobacco products in the United Kingdom. London, Department of Health. (*http://www.dh.gov.uk/ en/Publicatios andstatistics/Publications/PublicationsPolicyAndGuidance/DH_4070810),* 1-29.

Department of Health (2006). Consultation on the Introduction of Picture Warnings on Tobacco Packs. Wellington House, Department of Health, Tobacco Policy Team, Health Improvement Directorate. *(http://www. dh.gov.uk/en/Consultations/-Responsestoconsultations/DH_077960)*

DeVellis R (1991). *Scale Development: Theory and Applications*. Thousand Oaks, CA, Sage Publications. *(http://www.amazon.com/gp/reader/0761926054/ref=sib_dp_pt#reader-link)*

Devlin E, Anderson S, Borland R, et al. (2006). Development of a research tool to monitor point-of-sale promotions. *Soc Mark Q*, XII(1):29-39.

Diaz FJ, Jane M, Salto E, et al. (2005). A brief measure of high nicotine dependence for busy clinicians and large epidemiological surveys. *Aust N Z J Psychiatry*, 39(3):161-168.

Dierker LC, Canino G, Merikangas KR (2006). Association between parental and individual psychiatric/substance use disorders and smoking stages among Puerto Rican adolescents. *Drug Alcohol Depend*, 84(2):144-153.

DiFranza JR (1995). The Effects of tobacco advertising on children. In: *Proceedings of the 9th World Conference on Health, 10-14 October 1994, Paris, France*. New York, Plenum Publishing Corporation, 87-90.

DiFranza JR, Richards JW, Paulman PM, et al. (1991). RJR Nabisco's cartoon camel promotes camel cigarettes to children. *JAMA*, 266(22):3149-3153.

DiFranza JR, Coleman M, St Cyr D (1999). A comparison of the advertising and accessibility of cigars, cigarettes, chewing to-

bacco, and loose tobacco**.** *Prev Med*, 9(5):321-326.

DiFranza JR, Savageau JA, Rigotti NA, et al. (2002a). Development of symptoms of tobacco dependence in youths: 30 month follow up data from the DANDY study. *Tob Control*, 11(3):228-235.

DiFranza JR, Savageau JA, Fletcher K, et al. (2002b). Measuring the loss of autonomy over nicotine use in adolescents: the DANDY (Development and Assessment of Nicotine Dependence in Youths) study. *Arch Pediatr Adolesc Med*, 156(4):397-403.

Dillman D (2007). *Mail and Internet Surveys: The Tailored Design Method*. New York, John Wiley.

Dillman DA, Christian LM (2005). Survey mode as a source of instability in responses across surveys. *Field Meth*, (17):30-52.

Ding YS, Yan XJ, Jain RB, et al. (2006). Determination of 14 polycyclic aromatic hydrocarbons in mainstream smoke from U.S. brand and non-U.S. brand cigarettes. *Environ Sci Technol*, 40(4):1133-1138.

Djordjevic MV, Stellman SD, Zang E (2000). Doses of nicotine and lung carcinogens delivered to cigarette smokers. *J Natl Cancer Inst*, 92(2):106-111.

Dohrenwend B, Colombotos J, Dohrenwend B (1968). Social distance and interviewer effects. *Public Opin Q*, 32:410-422.

Dolcini MM, Adler NE, Ginsberg D (1996). Factors influencing agreement between self-reports and biological measures of smoking among adolescents. *J Res Adolesc*, 6(4):515-542.

Dolcini MM, Adler NE, Lee P, et al. (2003). An assessment of the validity of adolescent self-reported smoking using three biological indicators. *Nicotine Tob Res*, 5(4):473-483.

Doll R (2004). Evolution of Knowledge of the Smoking Epidemic. In: Boyle P, Gray N, Henningfield J, et al., eds., *Tobacco Science, Policy and Public Health.* Oxford, Oxford University Press, 3-16.

Doll R, Hill AB (1950). Smoking and carcinoma of the lung; preliminary report. *Br Med J*, 2(4682):739-748.

Doll R, Hill AB (1954). The mortality of doctors in relation to their smoking habits; a preliminary report. *Br Med J*, 1(4877): 1451-1455.

Donovan RJ, Jancey J, Jones S (2002). Tobacco point of sale advertising increases positive brand user imagery. *Tob Control*, 11(3):191-194.

Dorfman L, Ervice J, Woodruff K (2002). *Voices for Change: A Taxonomy of Public Communications Campaigns and their Evaluation Challenges*. Berkeley,CA, Berkeley Media Studios Group.

Dresler CM, Leon ME, Straif K, et al. (2006). Reversal of risk upon quitting smoking. *Lancet*, 368(9533):348-349.

Driesener C, Romaniuk J (2006). Comparing methods of brand image measurement. *Int J Mark Res*, 48(6):681-698.

Duncan GJ, Kalton G (1987). Issues of design and analysis of surveys across time. *Int Stat Rev* 55(1):97-117.

Durrant R, Wakefield M, McLeod K, et al. (2003). Tobacco in the news: an analysis of newspaper coverage of tobacco issues in Australia, 2001. *Tob Control*, 2(Suppl 2):ii75-ii81.

Eaton DK, Lowry R, Brener ND, et al. (2004). Passive versus active parental permission in school-based survey research: does the type of permission affect prevalence estimates of risk behaviors? *Eval Rev*, 28(6):564-577.

Ebbert JO, Patten CA, Schroeder DR (2006). The Fagerstrom Test for nicotine dependence-smokeless tobacco (FTND-ST). *Addict Behav*, 31(9): 1716-1721.

Edwards D, Riordan S (1994). Learned resourcefulness in black and white South African university students. *J Soc Psychol*, 134(5):665-675.

Edwards G, Gross MM (1976). Alcohol dependence: provisional description of a clinical syndrome. *Br Med J*, 1(6017): 1058-1061.

Eisner MD, Smith AK, Blanc PD (1998). Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA*, 280(22): 1909-1914.

Ellingsen DG, Fladseth G, Daae HL, et al. (2006). Airborne exposure and biological monitoring of bar and restaurant workers before and after the introduction of a smoking ban. *J Environ Monit*, 8(3):362-368.

Ellingson J, Sackett P, Hoagh LM (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *J Appl Psychol*, 84:155-166.

Elliot & Shanahan Research (2002). *Developmental Research for New Australian Health Warnings on Tobacco Products Stage 1. (http://www. health.gov.au/tobacco)*

Engelen M, Farrelly M, Hyland A (2006). *The Health and Economic Impact of New York's Clean Indoor Air Act*. Tobacco Control Program, Albany, New York State Department of Health.

Environics Research Group (2000). *Testing new health warning messages for cigarette packages: a summary of three phases of focus group research final report*. A report prepared by Environics Research Group Limited for Health Canada. Toronto, Ontario Canada, Environics Research Group Limited.

Environics Research Group (2003). *Summary report of four focus groups in Toronto & Montreal on awareness and understanding on toxic emissions information on tobacco packaging*. A report prepared by Environics Research Group Limited for Health Canada. Toronto, Ontario Canada, Environics Research Group Limited.

The European Public Health Alliance (2007). The European Public Health Alliance: EU Action Plan on Drugs. (*http://www epha org/a/1687)*.

Eremenco SL, Cella D, Arnold BJ (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof*, 28(2):212-232.

Ericsson K, Simon H (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MIT Press.

ESPAD. (2007). The European School Survey Project on Alcohol and Other Drugs. (*http://www.Espad.org)*.

Etter JF (2005). A comparison of the content-, construct- and predictive-validity of the cigarette dependence scale and the Fagerstrom test for nicotine dependence. *Drug Alcohol Depend*, 77(3):259-268.

Etter JF, Perneger TV (1999). Associations between the stages of change and the pros and cons of smoking in a longitudinal study of swiss smokers. *Addict Behav*, 24(3):419-424.

Etter JF, Perneger TV, Ronchi A (1997). Distributions of smokers by stage: international comparison and association with smoking prevalence. *Prev Med*, 26(4):580-585.

Etter JF, Duc TV, Perneger TV (1999). Validity of the Fagerstrom test for nicotine dependence and of the Heaviness of Smoking Index among relatively light smokers. *Addiction*, 94(2):269-281.

Etter JF, Humair JP, Bergman MM, et al. (2000a). Development and validation of the Attitudes Towards Smoking Scale (ATS-18). *Addiction*, 95(4):613-625.

Etter JF, Bergman MM, Humair JP, et al. (2000b). Development and validation of a scale measuring self-efficacy of current and former smokers. *Addiction*, 95(6):901-913.

Etter JF, Pelissolo A, Pomerleau C, et al. (2003a). Associations between smoking and heritable temperament traits. *Nicotine Tob Res*, 5(3):401-409.

Etter JF, Le Houezec J, Perneger TV (2003b). A self-administered questionnaire to measure dependence on cigarettes: the cigarette dependence scale. *Neuropsychopharmacology*, 28(2):359-370.

Etter JF, Kozlowski LT, Perneger TV (2003c). What smokers believe about light and ultralight cigarettes. *Prev Med*, 36(1):92-98.

European Network of Quitlines (2004). Guide to Best Practice. London, UK *(http://www.enqonline.org/public/aboutus.php)*

Evans NJ, Gilpin E, Pierce JP, et al. (1992). Occasional smoking among adults: evidence from the California Tobacco Survey. *Tob Control*, 1(3):169-175.

Evans N, Farkas A, Gilpin E, et al. (1995). Influence of tobacco marketing and exposure to smokers on adolescent susceptibility to smoking. *J Natl Cancer Inst*, 87(20):1538-1545.

Everett Jones S, Sharp DJ, Husten CG, et al. (2002). Cigarette acquisition and proof of age among U.S. high school students who smoke. *Tob Control*, 11(1):20-25.

Ezzati M, Lopez AD (2003). Estimates of global mortality attributable to smoking in 2000. *Lancet*, 362(9387):847-852.

Ezzati M, Lopez AD (2004). Regional, disease specific patterns of smoking-attributable mortality in 2000. *Tob Control*, 13(4):388-395.

Fagerstrom KO (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict Behav*, 3(3-4):235-241.

Farrelly MC, Bray JW, Pechacek T, et al. (2001). Response by adults to increases in cigarette prices by sociodemographic characteristics. *South Econ J*, 68(1):156-165.

Farrelly MC, Healton CG, Davis KC, et al. (2002). Getting to the truth: evaluating national tobacco countermarketing campaigns. *Am J Public Health*, 92(6):901-907.

Farrelly MC, Pechacek TF, Chaloupka FJ (2003a). The impact of tobacco control program expenditures on aggregate cigarette sales: 1981-2000. *J Health Econ*, 22(5):843-859.

Farrelly MC, Niederdeppe J, Yarsevich J (2003b). Youth tobacco prevention mass media campaigns: past, present, and future directions. *Tob Control*, 12(Suppl 1):i35-i47.

Farrelly MC, Nonnemaker JM, Chou R, et al. (2005a). Changes in hospitality workers' exposure to secondhand smoke following the implementation of New York's smoke-free law. *Tob Control*, 14(4):236-241.

Farrelly MC, Davis KC, Haviland ML, et al. (2005b). Evidence of a dose-response relationship between "truth" antismoking ads and youth smoking prevalence. *Am J Public Health*, 95(3):425-431.

Feighery E, Borzekowski DL, Schooler C, et al. (1998). Seeing, wanting, owning: the relationship between receptivity to tobacco marketing and smoking susceptibility in young people. *Tob Control*, 7(2):123-128.

Feighery EC, Henriksen L, Wang Y, et al. (2006). An evaluation of four measures of adolescents' exposure to cigarette marketing in stores. *Nicotine Tob Res*, 8(6):751-759.

Ferguson J, Bauld L, Chesterman J, et al. (2005). The English smoking treatment services: one-year outcomes. *Addiction*, 100(Suppl 2):59-69.

Ferguson JA, Patten CA, Schroeder DR, et al. (2003). Predictors of 6-month tobacco abstinence among 1224 cigarette smokers treated for nicotine dependence. *Addict Behav*, 28(7):1203-1218.

Ferris Wayne G, Connolly GN, Henningfield JE (2006). Brand differences of free-base nicotine delivery in cigarette smoke: the view of the tobacco industry documents. *Tob Control*, 15(3):189-198.

Fichtenberg CM, Glantz SA (2002a). Effect of smoke-free workplaces on smoking behaviour: systematic review. *Br Med J*, 325(7357):188.

Fichtenberg CM, Glantz SA (2002b). Youth access interventions do not affect youth smoking. *Pediatrics*, 109(6):1088-1092.

Fidler JA, Wardle J, Brodersen NH, et al. (2006). Vulnerability to smoking after trying a single cigarette can lie dormant for three years or more. *Tob Control*, 15(3):205-209.

Fielding JE, Briss PA, Carande-Kulis VG, et al. (2005). Tobacco. In: Zaza S, Briss PA, Harris KW, eds., *The Guide to Community Preventive Services: What Works to Promote Health?* New York, Oxford University Press.

Finkel S, Guterbock T, Borg M (1991). Race-of-interviewer effects in a presidential poll: Virginia 1989. *Public Opin Q*, 55:313-330.

Fiore MC, Novotny TE, Pierce JP, et al. (1990). Methods used to quit smoking in the United States. Do cessation programs help? *JAMA*, 263(20):2760-2765.

Fiore MC, Bailey WC, Cohen SJ, et al. (1996). *Smoking Cessation Clinical Practice Guideline No. 18*. Rockville, MD, Agency for Health Care Policy and Research, US Department of Health and Human Services. Publication No. 96-0692. *(http://www.surgeongeneral.gov/tobacco/default.htm)*

Fiore MC, Bailey WC, Cohen SJ (2000). *Treating Tobacco Use and Dependence*. Clinical Practice Guideline. Rockville, MD, US Department of Health and Human Services.

Fischer S, Spiegelhalder B, Preussmann R (1989a). Preformed tobacco-specific nitrosamines in tobacco: role of nitrate and influence of tobacco type. *Carcinogenesis*, 10(8):1511-1517.

Fischer S, Spiegelhalder B, Eisenbarth J, et al. (1990). Investigations on the origin of tobacco-specific nitrosamines in mainstream smoke of cigarettes. *Carcinogenesis*, 11(5):723-730.

Fischer PM, Richards JW Jr, Berman EJ, et al. (1989b). Recall and eye tracking study of adolescents viewing tobacco advertisements. *JAMA*, 261(1):84-89.

Fisher R, Katz J (2000). Social desirability bias and the validity of self-reported values. *Psychol Mark*, 17: 105-120.

Fiske AP, Kitayama S, Markus HR, et al. (1998). The cultural matrix of social psychology. In: Gilbert D, Fiske S, Lindzey G, eds., *Handbook of Social Psychology*. Boston, McGraw-Hill, 915-981.

Flay BR, Biglan A, Boruch RF, et al. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prev Sci*, 6(3):151-175.

Foddy W (1993). *Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research*. Cambridge, Cambridge University Press.

Fong GT, Cameron AJR, Brown KS, et al. (2002). Effect of the Canadian Graphic Warning Labels among High School Students: A Quasi-Experimental Longitudinal Survey. In: *Proceedings of the 2002 National Conference on Tobacco or Health,* 19-21 Nov 2002, *San Francisco, CA*. (*http://ncth.confex.com/ncth/2002/techprogram/paper_6056.htm*)

Fong GT, Cummings KM, Borland R, et al. (2006a). The conceptual framework of the International Tobacco Control (ITC) Policy Evaluation Project. *Tob Control*, 15(Suppl 3):iii3-11.

Fong GT, Hyland A, Borland R, et al. (2006b). Reductions in tobacco smoke pollution and increases in support for smoke-free public places following the implementation of comprehensive smoke-free workplace legislation in the Republic of Ireland: findings from the ITC Ireland/UK Survey. *Tob Control*, 15(Suppl 3):iii51-iii58.

Fontaine J (2003). Multidimensional Scaling. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-cultural Survey Methods.* Hoboken, John Wiley & Sons: 235-246.

Forsyth B, Rothgeb JM, Willis GB (2004). Does pretesting make a difference? An experimental test. In: Rothgeb JM, Couper MP, Lessler JT, et al., eds., *Methods for Testing and Evaluating Survey Questionnaires.* Hoboken, John Wiley & Sons, 525-546.

Foss OP, Haug K, Hesla PE, et al. (1998). [Can we rely on self-reported smoking habits?]. *Tidsskr Nor Laegeforen*, 118:2165-2168.

Foulds J, Ramstrom L, Burke M, et al. (2003). Effect of smokeless tobacco (snus) on smoking and public health in Sweden. *Tob Control*, 12(4):349-359.

Fowler F (1995). *Improving Survey Questions: Design and Evalua-tion*.Thousand Oaks, CA, Sage Publications. *(http:// www.amazon. com/gp/ reader/-0803945833/ref=sib_dp_pt#reader-link)*

Fowler FJ (2001). *Survey Research Methods.* Thousand Oaks, CA, Sage Publications. *(http://www.amazon. com/gp/reader/0761921915/ref=sib_dp_pt#reader-link)*

Fowler FJ, Stringfellow V (2001). Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols. *J Drug Issues*, 31(3):643-664.

Fowles J, Dybing E (2003). Application of toxicological risk assessment principles to the chemical constituents of cigarette smoke. *Tob Control*, 12(4):424-430.

Framework Convention Alliance (2007). *Civil Society Monitoring of the framework convention on tobacco control: 2007 Status Report.* Jategaonkar N, ed., Geneva, Framework Convention Alliance.

Freimuth V, Cole G, Kirby SD (2001). Issues in evaluating mass-media health communication campaigns. In: *Evaluation in Health Promotion: Principles and Perspectives*. Copenhagen, WHO Regional Office for Europe, 475-492.

Gajalakshmi CK, Jha P, Ranson K, et al. (2000). Global patterns of smoking and

smoking attributable mortality. In: Jha P, Chaloupka FJ, eds., *Tobacco Control in Developing Countries*. Oxford, Oxford University Press, 11-39.

Galbraith JW, Kaiserman M (1997). Taxation, smuggling and demand for cigarettes in Canada: evidence from time-series data. *J Health Econ*, 16(3):287-301.

Gambone MA (1998). Challenges of measurement in community change initiatives. In: Fulbright-Anderson K, Kubisch AC, Connell JP, eds., *New Approaches to Evaluating Community Initiatives*. Washington, D.C., The Aspen Institute of Human Studies, (*http://www.aspeninstitute.org/site/c.huLWJeMRKpH/b.613723/k.B03C/Challenges_of_Measurement_in_Community_Change_Initiatives.htm*).

GATS (2007). The Global Adult Tobacco Survey. *(http://www.cdc.gov/tobacco/global/ surveys)*

Gelernter J, Panhuysen C, Weiss R, et al. (2007). Genomewide linkage scan for nicotine dependence: identification of a chromosome 5 risk locus. *Biol Psychiatry*, 61(1):119-126.

Gfroerer J (1985). Influence of privacy on self-reported drug use by youths. *NIDA Res Monogr*, 57:22-30.

Gfroerer J, Wright D, Kopstein A (1997). Prevalence of youth substance use: the impact of methodological differences between two national surveys. *Drug Alcohol Depend*, 47(1):19-30.

Gillespie J, Milne K, Wilson N (2005). Secondhand smoke in New Zealand homes and cars: exposure, attitudes, and behaviours in 2004. *N Z Med J*, 118 (1227): U1782.

Gillmore J (2000). *Report on smoking prevalence in Canada, 1985 to 1999*. Ottawa, Health Statistics Division Statistics Canada. (*http://www.statcan. ca/english/research/82F0077XIE/82F0077XIE.pdf)*

Gilmore AB, McKee M (2004). Tobacco and transition: an overview of industry investments, impact and influence in the former Soviet Union. *Tob Control*, 13(2):136-142.

Gilmore AB, McKee M (2005). Exploring the impact of foreign direct investment on tobacco consumption in the former Soviet Union. *Tob Control*, 14(1):13-21.

Gilpin E, Pierce JP (1994). Measuring smoking cessation: problems with recall in the 1990 California Tobacco Survey. *Cancer Epidemiol Biomarkers Prev*, 3(7):613-617.

Gilpin EA, Pierce JP, Cavin SW, et al. (1994). Estimates of population smoking prevalence: self- versus proxy-reports of smoking status. *Am J Public Health*, 84(10):1576-1579.

Gilpin EA, Pierce JP, Rosbrook B (1997). Are adolescents receptive to current sales promotion practices of the tobacco industry? *Prev Med*, 26(1):14-21.

Giovino GA (2002). Epidemiology of tobacco use in the United States. *Oncogene*, 21(48):7326-7340.

Giovino GA, Shelton DM, Schooley MW (1993). Trends in cigarette smoking cessation in the United States. *Tob Control*, 2:S3-S10.

Giovino GA, Tomar SL, Reddy MN, et al. (2000). *Attitudes, Knowledge and Beliefs about Low-yield Cigarettes among Adolescents and Adults*. Oxford, Oxford University Press.

Glantz SA, Smith LR (1994). The effect of ordinances requiring smoke-free restaurants on restaurant sales. *Am J Public Health*, 84(7):1081-1085.

Glantz SA, Mandel LL (2005). Since school-based tobacco prevention programs do not work, what should we do? *J Adolesc Health*, 36(3):157-159.

Glassman AH, Helzer JE, Covey LS, et al. (1990). Smoking, smoking cessation, and major depression. *JAMA*, 264(12):1546-1549.

Global Tobacco Surveillance System Collaborating Group (GTSS) (2005). Global Tobacco Surveillance System (GTSS): purpose, production, and potential. *J Sch Health*, 75(1):15-24.

Global Youth Tobacco Survey Collaborative Group (2002) Tobacco use among youth: a cross country comparison. *Tob Control*, 11(3):252-270.

Godeau E, Rahav G, Hublet A (2004). Tobacco Smoking. In: Currie C, World Health Organization, Regional Office for Europe, eds., *Young People's Health in Context: Health Behaviour in School-aged Children (HBSC) Study: International Report from the 2001/2002 survey.* Copenhagen, Denmark, World Health Organization, Regional Office for Europe, 63-72.

Godfrey C, Parrott S, Coleman T, et al. (2005). The cost-effectiveness of the English smoking treatment services: evidence from practice. *Addiction*, 100(Suppl 2):70-83.

Goldberg ME, Liefield J, Kindra G, et al. (1995). *When packages can't speak: Possible impacts of plain and generic packaging of tobacco products*. Expert Panel Report. Ottawa, Health Canada

Gori GB (1990). Consumer perception of cigarette yields: is the message relevant? *Regul Toxicol Pharmacol*, 12(1):64-68.

Gospodinov N, Irvine IJ (2004). Global health warnings on tobacco packaging: Evidence from the Canadian experiment. *B E J Econ Anal Policy*, 4 (Issue 1).

Grant BF, Hasin DS, Chou SP, et al. (2004). Nicotine dependence and psychiatric disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Arch Gen Psychiatry*, 61(11):1107-1115.

Gray N, Boyle P (2002). Regulation of cigarette emissions. *Ann Oncol*, 13(1):19-21.

Gray N, Zaridze D, Robertson C, et al. (2000). Variation within global cigarette brands in tar, nicotine, and certain nitrosamines: analytic study. *Tob Control*, 9(3):351.

Green CR, Rodgman A (1996). The tobacco chemists' research conference: a half century forum for advances in analytical methodology of tobacco and its products. *Recent Adv Tob Sci*, 131.

Greenleaf EA (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *J Mark Res*, 19:176-188.

Gritz ER, Baer-Weiss V, Benowitz NL, et al. (1981). Plasma nicotine and cotinine concentrations in habitual smokeless tobacco users. *Clin Pharmacol Ther*, 30(2):201-209.

Gritz ER, Schacherer C, Koehly L, et al. (1999). Smoking withdrawal and relapse in head and neck cancer patients. *Head Neck*, 21(5):420-427.

Grosh ME, Glewwe P (1998). The World Bank's living standards measurement study household surveys. *J Econ Perspect*, 12(1):187-196.

Groves RM (2001). *Survey Nonresponse*. New York, John Wiley and Sons.

GSHS (2007). The Global School-Based Student Health Survey. (*http://www.who.int/chp/gshs/en*; *http://www.cdc.gov/gshs*).

Guillemin F, Bombardier C, Beaton D (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*, 46(12):1417-1432.

Guindon GE, Boisclair D (2003). *Past, Current and Future Trends in Tobacco Use*. Washington, DC, World Bank.

Gulliver SB, Hughes JR, Solomon LJ, et al. (1995). An investigation of self-efficacy, partner support and daily stresses as predictors of relapse to smoking in self-quitters. *Addiction*, 90(6):767-772.

Guttman N, Peleg H (2003). Public preferences for an attribution to government or to medical research versus unattributed messages in cigarette warning labels in Israel. *Health Commun*, 15(1):1-25.

GYTS (2007). The Global Youth Tobacco Survey. (*http://www.cdc.gov/tobacco/global/surveys)*.

Haddock CK, Lando H, Klesges RC, et al. (1999). A study of the psychometric and predictive properties of the Fagerström Test for nicotine dependence in a population of young smokers. *Nicotine Tob Res,* 1(1):59-66.

Hambleton RK, Merenda PF, Spielberger CD (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, Erlbaum Assoc.

Hamilton WL, Norton G, Ouellette TK, et al. (2004). Smokers' responses to advertisements for regular and light cigarettes and potential reduced-exposure tobacco products. *Nicotine Tob Res*, 6(Suppl 3):S353-S362.

Hammond D, Fong GT, McDonald PW, et al. (2003). Impact of the graphic Canadian warning labels on adult smoking behaviour. *Tob Control*, 12(4):391-395.

Hammond D, Fong GT, McDonald PW, et al. (2004a). Graphic Canadian cigarette warning labels and adverse outcomes: evidence from Canadian smokers. *Am J Public Health*, 94(8):1442-1445.

Hammond D, McDonald PW, Fong GT, et al. (2004b). The impact of cigarette warning labels and smoke-free bylaws on smoking cessation: evidence from former smokers. *Can J Public Health*, 95(3):201-204.

Hammond D, Fong GT, McNeill A, et al. (2006a). Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15(Suppl 3):iii19-iii25.

Hammond D, Fong GT, Cummings KM, et al. (2006b). Cigarette yields and human exposure: a comparison of alternative testing regimens. *Cancer Epidemiol Biomarkers Prev*, 15(8):1495-1501.

Hammond D, Fong GT, Borland R, et al. (2007a). Text and graphic warnings on cigarette packages: findings from the international tobacco control four country study. *Am J Prev Med*, 32(3):202-209.

Hammond D, Wiebel F, Kozlowski LT, et al. (2007b). Revising the machine smoking regime for cigarette emissions: implications for tobacco control policy. *Tob Control*, 16(1):8-14.

Hammond SK (1999). Exposure of U.S. workers to environmental tobacco smoke. *Environ Health Perspect*, 107(Suppl 2):329-340.

Hammond SK, Coghlin J, Gann PH, et al. (1993). Relationship between environmental tobacco smoke exposure and carcinogen-hemoglobin adduct levels in nonsmokers. *J Natl Cancer Inst*, 85(6):474-478.

Hanley JA, Negassa A, Edwardes MD, et al. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol*, 157(4):364-375.

Harkness JA (2003). Questionnaire Translation. In: Harkness JA, Van de Vijver F, Mohler PP, eds., *Cross-Cultural Survey Methods*. Hoboken, John Wiley & Sons, 35-56.

Harkness JA, Schoua-Glusberg A (1998). Questionnaires in Translation. In: Harkness JA, ed., *Cross-Cultural Survey Equivalence.* Mannheim, ZUMA, 87-126.

Harkness JA, Blom A (2006). *Round 3 ESS Translation Guidelines*. Mannheim, ZUMA.

Harkness JA, Van de Vijver FJR, Mohler PP (2003a). *Cross-Cultural Survey Methods.* Hoboken, John Wiley & Sons.

Harkness JA, Van de Vijver FJR, Johnson TP (2003b). Questionnaire Design in Comparative Research. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Survey Methods.* Hoboken, John Wiley & Sons, 19-34.

Harkness JA, Pennell BA, Schoua-Glusberg A (2004). Survey Questionnaire Translation and Assessment. In: Pressler S, Rothgeb JM, Couper MP, et al., eds.,

Methods for Testing and Evaluating Survey Questionnaires. Hoboken, John Wiley & Sons, 453-473.

Harold C, Pillsbury Jr (1996). Review of the Federal Trade Commission Method for Determining Cigarette Tar and Nicotine Yield. In: Shopland DR, ed. *The FTC Cigarette Test Method for Determining Tar, Nicotine, and Carbone Monoxide Yields of U.S. Cigarettes.* Report of the NCI Expert Committee. Smoking and Tobacco Control Monograph No. 7. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub No. 96-4028; 9-14.*(http://cancercontrol.cancer.gov/tcrb/monographs/7/m7_2.pdf)*

Harris JE (2001). Smoke yields of tobacco-specific nitrosamines in relation to FTC tar level and cigarette manufacturer: analysis of the Massachusetts Benchmark Study. *Public Health Rep*, 116(4):336-343.

Hart P (1996). Product placement comes of age in Canada. *Marketing*, 101:20.

Hastings G, MacFadyen L (2000). A day in the life of an advertising man: review of internal documents from the UK tobacco industry's principal advertising agencies. *Br Med J*, 321(7257):366-371.

Hatsukami D, Anton D, Keenan R, et al. (1992). Smokeless tobacco abstinence effects and nicotine gum dose. *Psychopharmacology (Berl)*, 106(1): 60-66.

Hatsukami DK, Severson HH (1999). Oral spit tobacco: addiction, prevention and treatment. *Nicotine Tob Res*, 1(1):21-44.

Hatsukami DK, Slade J, Benowitz NL, et al. (2002). Reducing tobacco harm: research challenges and issues. *Nicotine Tob Res*, 4(Suppl 2):S89-101.

Hatsukami DK, Lemmonds C, Zhang Y, et al. (2004). Evaluation of carcinogen exposure in people who used "reduced exposure" tobacco products. *J Natl Cancer Inst*, 96(11):844-852.

Hatsukami DK, Giovino GA, Eissenberg T, et al. (2005). Methods to assess potential reduced exposure products. *Nicotine Tob Res*, 7(6):827-844.

Hatsukami DK, Benowitz NL, Rennard SI, et al. (2006). Biomarkers to assess the utility of potential reduced exposure tobacco products. *Nicotine Tob Res*, 8(4):600-622.

Hatziandreu EJ, Pierce JP, Fiore MC, et al. (1989). The reliability of self-reported cigarette consumption in the United States. *Am J Public Health*, 79(8):1020-1023.

HBSC (2007). *The Health Behaviour in School-Aged Children Survey*. (*http://www.Hbsc.org*)

Health Canada (1999a). *Preparation of Cigarettes from Packaged Leaf Tobacco for Testing*. Ottawa, Canada, Health Canada. *(http://www-lib/doc-lib/TOB200703/Health%20Canada% 201999.pdf)*

Health Canada (1999b). *Determination of "Tar", Nicotine and Carbon Monoxide in Mainstream Tobacco Smoke, Official Method T-115*. (*http://www hc-sc gc ca/hl-vs/alt_formats/hecs-sesc/pdf/t obac-tabac/legislation/reg/indust/ method/main-principal/nicotine/nicotine_e pdf*).

Health Canada (2001). *Baseline surveys: the health effects of tobacco and health warning messages on cigarette packages: final report*. Ottawa, Health Canada, Environics Research Group.

Health Canada (2003). *Toxics information on cigarette packaging: results of a survey of smokers*. Ottawa, Environics Research Group.

Health Canada (2005). *Baseline Surveys: The Health Effects of Tobacco and Health Warning Messages on Cigarette Packages: Wave 9 Surveys*. Ottawa, Environics Research Group.

Health Canada (2006). *Qualitative Testing of Health Warning Messages*. Ottawa, Health Canada, CREATEC.

Heatherton TF, Kozlowski LT, Frecker RC, et al. (1989). Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *Br J Addict*, 84(7):791-799.

Heatherton TF, Kozlowski LT, Frecker RC, et al. (1991). The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict*, 86(9):1119-1127.

Hecht SS (1998). Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem Res Toxicol*, 11(6):559-603.

Hecht SS (1999). Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst*, 91(14):1194-1210.

Hecht SS (2002). Human urinary carcinogen metabolites: biomarkers for investigating tobacco and cancer. *Carcinogenesis*, 23(6):907-922.

Hecht SS (2004). Carcinogen derived biomarkers: applications in studies of human exposure to secondhand tobacco smoke. *Tob Control*, 13(Suppl 1):i48-i56.

Hecht SS, Carmella SG, Ye M, et al. (2002). Quantitation of metabolites of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone after cessation of smokeless tobacco use. *Cancer Res*, 2(1):129-134.

Hecht SS, Murphy SE, Carmella SG, et al. (2004). Effects of reduced cigarette smoking on the uptake of a tobacco-specific lung carcinogen. *J Natl Cancer Inst*, 96(2):107-115.

Henderson B (2000). Wear out: An empirical investigation of advertising wear-in and wear-out. *J Advert Res*, 6:95-100.

Henkel V, Mergl R, Kohnen R, et al. (2003). Identifying depression in primary care: a comparison of different methods in a prospective cohort study. *Br Med J*, 326(7382):200-201.

Henkel V, Mergl R, Kohnen R, et al. (2004a). Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. *Gen Hosp Psychiatry*, 26(3):190-198.

Henkel V, Mergl R, Coyne JC, et al. (2004b). Screening for depression in primary care: will one or two items suffice? *Eur Arch Psychiatry Clin Neurosci*, 254(4):215-223.

Henley SJ, Connell CJ, Richter P, et al. (2007). Tobacco-related disease mortality among men who switched from cigarettes to spit tobacco**.** *Tob Control*, 16(1):22-28.

Henningfield JE, Fant RV, Tomar SL (1997). Smokeless tobacco: an addicting drug. *Adv Dent Res*, 11(3):330-335.

Henningfield J, Pankow J, Garrett B (2004). Ammonia and other chemical base tobacco additives and cigarette nicotine delivery: issues and research needs**.** *Nicotine Tob Res*, 6(2):199-205.

Henry GT, Gordon CS (2001). Tracking issue attention: Specifying the dynamics of the public agenda. *Public Opin Q*, 65:157-177.

Hersey JC, Niederdeppe J, Ng SW, et al. (2005). How state counter-industry campaigns help prime perceptions of tobacco industry practices to promote reductions in youth smoking. *Tob Control*, 14(6):377-383.

Hibell B (2004). *The ESPAD report 2003 alcohol and other drug use among students in 35 European countries*. Stockholm, Swedish Council for Information on Alcohol and other Drugs.

Hill AB (1965). The environment and disease: association or causation? *Proc R Soc Med*, 58:295-300.

Hill D (1988). New cigarette-packet warnings: are they getting through? *Med J Aust*, 148(9):478-480.

HM Customs and Excise (2001). *Measuring indirect tax fraud*. London, HM Customs and Excise. *(http://www. hmce. gov.uk)*

Ho SY, Lam TH, Chung SF, et al. (2007). Cross-sectional and prospective associations between passive smoking and respiratory symptoms at the workplace. *Ann Epidemiol*, 17(2):126-131.

Hoch E, Muehlig S, Hofler M, et al. (2004). How prevalent is smoking and nicotine dependence in primary care in Germany? *Addiction*, 99(12):1586-1598.

Hoek J, Gendall P, Stockdale M (1993). Some effects of tobacco sponsorship advertisements on young males. *Int J Advert*, 12:13-28.

Hoffmann D, Brunnemann KD (1982). Pyrolytic origins of major gas phase constituents of gigarette smoke. *Recent Adv Tob Sci*, 8:103.

Hoffmann D, Hoffmann I (1997). The changing cigarette, 1950-1995. *J Toxicol Environ Health*, 50(4):307-364.

Hoffmann D, Hoffmann I (2001). The Changing Cigarette: Chemical Studies and Bioassays. In: Shopland DR, ed. *Risks Associated with Smoking Cigarettes with Low Machine-Measured Yields of Tar and Nicotine*. Smoking and Tobacco Control Monograph No.13. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub. No.02-5047; 159-191. *(http://cancercontrol. cancer.gov/tcrb/monographs/13/m13_5.pdf)*

Hoffmann D, Djordjevic MV, Fan J, et al. (1995). Five leading U.S. commercial brands of moist snuff in 1994: assessment of carcinogenic N-nitrosamines. *J Natl Cancer Inst*, 87(24):1862-1869.

Hong T, Cody MJ (2002). Presence of protobacco messages on the web. *J Health Commun*, 7:273-307.

Hopkins DP, Briss PA, Ricard CJ, et al. (2001). Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. *Am J Prev Med*, 20(2 Suppl 3.3):16-66.

Horm J, Cynamon M, Thornberry O (1996). The Influence of Parental Presence on the Reporting of Sensitive Behaviors by Youth. In: Warnecke R, ed., *Health Survey Research Methods: Conference Proceedings*. Washington, D.C., Department of Health and Human Services, 141-145.

Hornik R (2002). *Public Health Communication: Evidence for Behavior Change*. Mahwah, NJ, Lawrence Erlbaum Associates, Inc. *(http://www. amazon.com/gp/reader/0805831770/ref=sib_dp_pt# reader-link)*

Hovell MF, Zakarian JM, Wahlgren DR, et al. (2000). Reducing children's exposure to environmental tobacco smoke: the empirical evidence and directions for future research. *Tob Control*, 9(Suppl 2):II40-II47.

Howard LA, Ahluwalia JS, Lin SK, et al. (2003). CYP2E1*1D regulatory polymorphism: association with alcohol and nicotine dependence. *Pharmacogenetics*, 13(6):321-328.

Hoyle RH (1995). *Structural equation modeling; concepts, issues, and applications*. Thousand Oaks, CA, Sage Publications.*(http://www.amazon. com/gp/reader/0803953186/ref=sib_dp_pt#reader-link)*

Hoyle RH, Stephenson MT, Palmgreen P, et al. (2002). Reliability and validity of a brief measure of sensation seeking. *Pers Individ Diffs*, 32(3):401-414.

Huang CL, Lin HH, Wang HH (2006). Psychometric evaluation of the Chinese version of the Fagerstrom Tolerance Questionnaire as a measure of cigarette dependence. *J Adv Nurs*, 55(5):596-603.

Hublet A, De Bacquer D, Valimaa R, et al. (2006). Smoking trends among adolescents from 1990 to 2002 in ten European countries and Canada. *BMC Public Health*, 6:280.

Huerta M, Chodick G, Balicer RD, et al. (2005). Reliability of self-reported smoking history and age at initial tobacco use. *Prev Med*, 41(2):646-650.

Hughes JR, Keely JP, Niaura RS, et al. (2003). Measures of abstinence in clinical trials: issues and recommendations. *Nicotine Tob Res*, 5(1):13-25.

Hughes JR, Oliveto AH, Riggs R, et al. (2004a). Concordance of different measures of nicotine dependence: two pilot studies. *Addict Behav*, 29(8):1527-1539.

Hughes JR, Benowitz N, Hatsukami D, et al. (2004b). Clarification of SRNT workgroup guidelines for measures in clinical trials of smoking cessation therapies. *Nicotine Tob Res*, 6(5):863-864.

Hughes JR, Keely JP, Fagerstrom KO, et al. (2005). Intentions to quit smoking change over short periods of time. *Addict Behav*, 30(4):653-662.

Hughes JR, Helzer JE, Lindberg SA (2006). Prevalence of DSM/ICD-defined nicotine dependence. *Drug Alcohol Depend*, 85(2):91-102.

Husten CG, McCarty MC, Giovino GA, et al. (1998). Intermittent smokers: a descriptive analysis of persons who have never smoked daily. *Am J Public Health*, 88(1):86-89.

Hyland A, Cummings KM (1999a). Consumer response to the New York City Smoke-Free Air Act. *J Public Health Manag Pract*, 5(1):28-36.

Hyland A, Cummings KM (1999b). Restaurateur reports of the economic impact of the New York City Smoke-Free Air Act. *J Public Health Manag Pract*, 5(1):37-42.

Hyland A, Cummings KM (1999c). Restaurant employment before and after the New York City Smoke-Free Air Act. *J Public Health Manag Pract*, 5(1):22-27.

Hyland A, Cummings KM, Wilson MP (1999a). Compliance with the New York City Smoke-Free Air Act. *J Public Health Manag Pract*, 5(1):43-52.

Hyland A, Cummings KM, Nauenberg E (1999b). Analysis of taxable sales receipts: was New York City's Smoke-Free Air Act bad for restaurant business? *J Public Health Manag Pract*, 5(1):14-21.

Hyland A, Li Q, Bauer JE, et al. (2004). Predictors of cessation in a cohort of current and former smokers followed over 13 years. *Nicotine Tob Res*, 6(Suppl 3):S363-S369.

Hyland A, Higbee C, Li Q, et al. (2005). Access to low-taxed cigarettes deters smoking cessation attempts. *Am J Public Health*, 95(6):994-995.

Hyland A, Borland R, Li Q, et al. (2006). Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15(Suppl 3):iii83-iii94.

Hyland A, Higbee C, Hassan L, et al. (2007). Does smoke-free Ireland have more smoking inside the home and less in pubs than the United Kingdom? Findings from the International Tobacco Control Policy Evaluation Project. *Eur J Public Health* (Epub 2007 Jun 14).

Hyland M, Birrell J (1979). Government health warnings and the "boomerang" effect. *Psychol Rep*, 44(2):643-647.

Hymowitz N, Sexton M, Ockene J, et al. (1991). Baseline factors associated with smoking cessation and relapse. MRFIT Research Group. *Prev Med*, 20(5):590-601.

Hymowitz N, Cummings KM, Hyland A, et al. (1997). Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tob Control*, 6(Suppl 2):S57-S62.

IARC (1986). Environmental carcinogens. Selected methods of analysis. Volume 8: Some metals: As, Be, Cd, Cr, Ni, Pb, Se, Zn. *IARC Sci Publ*, (71):129-138.

IARC (2004). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 83, Tobacco Smoke and Involuntary Smoking*. Lyon, IARC Press.

IARC (2007a). *IARC Handbooks of Cancer Prevention. Tobacco Control. Volume 11: Reversal of Risk After Quitting Smoking.* Lyon, IARC Press.

IARC (2007b). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 89. Smokeless Tobacco Products*. Lyon, IARC Press.

Institute of Medicine (1988). *The Future of Public Health*. Washington, D.C., National Academy Press.

Institute of Medicine (2001). *Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction*. Washington, D.C., National Academy Press.

International Programme of Chemical Safety, the United Nations Environment Programme the International Labour Organisation World Health Organization. (1993). *Biomarkers and Risk Assessment: Concepts and Principles*. Geneva, World Health Organization.

ITC (2007). *The International Tobacco Control Policy Evaluation Project*. (*http://www.itcproject.org*).

Jackson R, Beaglehole R (1985). Secular trends in underreporting of cigarette consumption. *Am J Epidemiol*, 122(2):341-344.

Jacob P, III, Hatsukami D, Severson H, et al. (2002). Anabasine and anatabine as biomarkers for tobacco use during nicotine replacement therapy. *Cancer Epidemiol Biomarkers Prev*, 11(12):1668-1673.

Janz NK, Becker MH (1984). The Health Belief Model: a decade later. *Health Educ Q*, 11(1):1-47.

Jenkins RA, Guerin MR, Tomkins BA (2000). *The Chemistry of Environmental Tobacco Smoke: Composition and Measurement*. Boca Raton, Lewis Publishers.

Jha P, Chaloupka FJ (1999). *Curbing the Epidemic. Governments and the Economics of Tobacco Control*. Washington, D.C., World Bank.

Jha P, Chaloupka FJ (2000). *Tobacco control in developing countries*. Oxford, Oxford University Press.

Jha P, Chaloupka F, Moore J, et al. (2006). Tobacco Addiction. In: Jamison DT, Breman JG, Measham AR, et al., eds., *Disease Control Priorities in Developing Countries*. Washington D.C., The Interna-

tional Bank for Reconstruction and Development, World Bank, 869-885.

John U, Meyer C, Hapke U, et al. (2003a). The Fagerstrom test for nicotine dependence in two adult population samples-potential influence of lifetime amount of tobacco smoked on the degree of dependence. *Drug Alcohol Depend*, 71(1):1-6.

John U, Meyer C, Rumpf HJ, et al. (2003b). Strength of the relationship between tobacco smoking, nicotine dependence and the severity of alcohol dependence syndrome criteria in a population-based sample. *Alcohol*, 38(6):606-612.

John U, Meyer C, Rumpf HJ, et al. (2004a). Smoking, nicotine dependence and psychiatric comorbidity—a population-based study including smoking cessation after three years. *Drug Alcohol Depend*, 76(3):287-295.

John U, Meyer C, Schumann A, et al. (2004b). A short form of the Fagerstrom Test for Nicotine Dependence and the Heaviness of Smoking Index in two adult population samples. *Addict Behav*, 29(6):1207-1212.

John U, Meyer C, Rumpf HJ, et al. (2005). Self-rated general health and psychiatric disorders in a general population sample. *Eur Psychiatry*, 20(3):223-228.

Johnson EO, Breslau N, Anthony JC (1996). The latent dimensionality of DIS/DSM-III-R nicotine dependence: exploratory analyses. *Addiction*, 91(4):583-588.

Johnson EO, Chase GA, Breslau N (2002). Persistence of cigarette smoking: familial liability and the role of nicotine dependence. *Addiction*, 97(8):1063-1070.

Johnson T, Fendrich M, Shaligram C, et al. (2000). An evaluation of the effects of interviewer characteristics in an RDD telephone survey of drug use. *J Drug Issues*, 30:77-101.

Johnson TP (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. In: Harkness JA, ed., *Cross-cultural Survey Equivalence*. Mannheim, ZUMA, 1-40. *(http://www-lib/doc-lib/TOB-200703/Johnson1998.pdf)*.

Johnson TP, Mott JA (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction*, 96(8):1187-1198.

Johnson TP, Van de Vijver FJR (2004). Social desireability in cross-cultural research. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Survey Methods*. Hoboken, John Wiley & Sons, 195-204.

Johnson TP, Cho YI, Holbrook AL, et al. (2006). Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Ann Epidemiol*, 16(9):661-668.

Johnston LD (2001). Changing Demographic Patterns of Adolescent Smoking over the Past 23 Years: National Trends from the Monitoring the Future Study. In: Burns DM, Amacher RH and Ruppert W, eds. *Changing Adolescent Smoking Prevalence: Where It Is and Why*. Smoking and Tobacco Control Monograph No. 14. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub. 02-5086; No. 9-34.

Johnston LD, O'Malley PM, Bachman JG, et al. (2006). *Monitoring the Future national survey results on drug use, 1975-2005. Volume I: Secondary school students*. Bethesda, MD, National Institute on Drug Abuse. NIH Publication No. 06-5883.

Joossens L (2003). Vietnam: smuggling adds value. *Tob Control*, 12(2):119-120.

Joossens L, Raw M (1995). Smuggling and cross border shopping of tobacco in Europe. *Br Med J*, 310(6991):1393-1397.

Joossens L, Raw M (1998). Cigarette smuggling in Europe: who really benefits? *Tob Control*, 7(1):66-71.

Joossens L, Raw M (2002). *Turning off the Tap. An Update on Cigarette Smuggling in the UK and Sweden with Recommendations to Control Smuggling*. London, Cancer Research UK.

Joossens L, Raw M (2006). The Tobacco Control Scale: a new scale to measure country activity. *Tob Control*, 15(3):247-253.

Joreskog KG, Sorbom D (1996). *LISREL 8: User's Reference Guide*. Chicago, Scientific Software International.

Joseph AM, Hecht SS, Murphy SE, et al. (2005). Relationships between cigarette consumption and biomarkers of tobacco toxin exposure. *Cancer Epidemiol Biomarkers Prev*, 14(12): 2963-2968.

Judge K, Bauld L, Chesterman J, et al. (2005). The English smoking treatment services: short-term outcomes. *Addiction*, 100(Suppl 2):46-58.

Kann L, Brener ND, Warren CW, et al. (2002). An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *J Adolesc Health*, 31(4):327-335.

Kawakami N, Takatsuka N, Inaba S, et al. (1999). Development of a screening questionnaire for tobacco/nicotine dependence according to ICD-10, DSM-III-R, and DSM-IV. *Addict Behav*, 24(2):155-166.

Keeler TE, Hu TW, Barnett PG, et al. (1993). Taxation, regulation, and addiction: a demand function for cigarettes based on time-series evidence. *J Health Econ*, 12(1):1-18.

Kendler KS, Neale MC, Sullivan P, et al. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol Med*, 29(2):299-308.

Kendrick JS, Zahniser SC, Miller N, et al. (1995). Integrating smoking cessation into routine public prenatal care: the Smoking Cessation in Pregnancy project. *Am J Public Health*, 85(2):217-222.

Kessler DA (1995). Nicotine addiction in young people. *N Engl J Med*, 333(3):186-189.

Kessler RC, Andrews G, Colpe LJ, et al. (2002). Short screening scales to monitor population prevalences and trends in non-

specific psychological distress. *Psychol Med*, 32(6):959-976.

King B, Borland R (2004). The "low-tar" strategy and the changing construction of Australian cigarettes. *Nicotine Tob Res*, 6(1):85-94.

Klesges RC, Debon M, Ray JW (1995). Are self-reports of smoking rate biased? Evidence from the Second National Health and Nutrition Examination Survey. *J Clin Epidemiol*, 48(10):1225-1233.

Kline RB (2005). *Principles and practice of structural equation modeling* . New York, Guilford Press. *(http://www.loc.gov/catdir/bios/guilford051/2004019784.html;http://www.loc.gov/catdir/description/guilford051/2004019784.html;http://www.loc.gov/catdir/toc/ecip0422/2004019784.html)*

Knowles E, Condon C (1999). Why people say "yes": A dual-process theory of acquiescence. *J Pers Soc Psychol*, 77:379-386.

Koenen KC, Hitsman B, Lyons MJ, et al. (2005). A twin registry study of the relationship between posttraumatic stress disorder and nicotine dependence in men. *Arch Gen Psychiatry*, 62(11):1258-1265.

Kopstein AN (2001). *Tobacco use in America: Findings from the 1999 National Household Survey on Drug Abuse*. Rockville, MD, US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies. *(http://www. oas.samhsa.gov/NHSDA/tobacco/toc.htm)*

Koval JJ, Aubut JAL, Pederson LL, et al. (2005). The potential effectiveness of warning labels on cigarette packages. The perceptions of young adult Canadians. *Can J Public Health*, 96(5):353-356.

Kozlowski LT, O'Connor RJ (2000). Official cigarette tar tests are misleading: use a two-stage, compensating test. *Lancet*, 355(9221):2159-2161.

Kozlowski LT, O'Connor RJ (2002). Cigarette filter ventilation is a defective design because of misleading taste, bigger puffs, and blocked vents. *Tob Control*, 11(Suppl 1):I40-I50.

Kozlowski LT, Frecker RC, Khouw V, et al. (1980). The misuse of 'less-hazardous' cigarettes and its detection: hole-blocking of ventilated filters. *Am J Public Health*, 70(11):1202-1203.

Kozlowski LT, Porter CQ, Orleans CT, et al. (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HSI. *Drug Alcohol Depend*, 34(3):211-216.

Kozlowski LT, Goldberg ME, Yost BA, et al. (1998b). Smokers' misperceptions of light and ultra-light cigarettes may keep them smoking. *Am J Prev Med*, 15(1):9-16.

Kozlowski LT, Mehta NY, Sweeney CT, et al. (1998a). Filter ventilation and nicotine content of tobacco in cigarettes from Canada, the United Kingdom, and the United States. *Tob Control*, 7(4):369-375.

Kozlowski LT, Goldberg ME, Yost BA (2000). Measuring smokers' perceptions of the health risks from smoking light cigarettes. *Am J Public Health*, 90(8):1318-1319.

Kozlowski LT, O'Connor RJ, Giovino GA, et al. (2006). Maximum yields might improve public health—if filter vents were banned: a lesson from the history of vented filters. *Tob Control*, 15(3):262-266.

Krugman DM, Fox RJ, Fletcher JE, et al. (1994). Do adolescents attend to warnings in cigarette advertising? An eye tracking approach. *J Advert Res*, 34:39-52.

Kuechler M. (1987). The Utility of Surveys for Cross-National Research. *Soc Sci Res*, 16:229-244.

Kyaing NN, Perucic AM, Rahman K (2005). *Study on poverty alleviation and tobacco control in Myanmar*. Economics of Tobacco Control Paper N°31,Washington DC, The International Bank for Reconstruction and Development, the World Bank. *(http://www.worldbank.org/tobacco)*

Laatikainen T, Vartiainen E, Puska P (1999). Comparing smoking and smoking cessation process in the Republic of Karelia, Russia and North Karelia, Finland. *J Epidemiol Commun Health*, 53(9):528-534.

Lam TH, Ho LM, Hedley AJ, et al. (2005). Secondhand smoke and respiratory ill health in current smokers. *Tob Control*, 14(5):307-314.

Lance PM, Akin JS, Dow WH, et al. (2004). Is cigarette smoking in poorer nations highly sensitive to price? Evidence from Russia and China. *J Health Econ*, 23(1):173-189.

Lantz PM, Jacobson PD, Warner KE, et al. (2000). Investing in youth tobacco control: a review of smoking prevention and control strategies. *Tob Control*, 9(1):47-63.

Larabie LC (2005). To what extent do smokers plan quit attempts? *Tob Control*, 14(6):425-428.

Lasser K, Boyd JW, Woolhandler S, et al. (2000). Smoking and mental illness: A population-based prevalence study. *JAMA*, 284(20):2606-2610.

Law MR, Wald NJ (2003). Environmental tobacco smoke and ischemic heart disease. *Prog Cardiovasc Dis*, 46(1):31-38.

Laws MB, Whitman J, Bowser DM, et al. (2002). Tobacco availability and point of sale marketing in demographically contrasting districts of Massachusetts. *Tob Control*, 11(Suppl 2):ii71-ii73.

Laxminarayan R, Deolalikar A (2004). Tobacco initiation, cessation, and change: evidence from Vietnam. *Health Econ*, 13(12):1191-1201.

Leaderer BP, Koutrakis P, Briggs SLK, et al. (1994). The mass concentration and elemental composition of indoor aerosols in Suffolk and Onondaga Counties, New York. *Indoor Air - Intl J Indoor Air Quality Climate*, 4(1):23-34.

Leading Edge Consultancy (1997). *American Blends NPD, Qualitative Debrief Pres-*

entation. *(http://www-lib/doc-lib/TOB 200703/Leading-%20Edge%20Consultancy%201997.pdf)*

Leatherdale ST (2005). Predictors of different cigarette access behaviours among occasional and regular smoking youth. *Can J Public Health*, 96(5):348-352.

Ledwith F (1984). Does sports sponsorship act as advertising to children? *Health Educ J*, 43:85-88.

Lee CK, Kwak YS, Yamamoto J, et al. (1990). Psychiatric epidemiology in Korea. Part I: Gender and age differences in Seoul. *J Nerv Ment Dis*, 178(4):242-246.

Lee EM, Malson JL, Waters AJ, et al. (2003). Smoking topography: reliability and validity in dependent smokers. *Nicotine Tob Res*, 5(5):673-679.

Lemmonds CA, Hecht SS, Jensen JA, et al. (2005). Smokeless tobacco topography and toxin exposure. *Nicotine Tob Res*, 7(3):469-474.

Lerman C, Tyndale R, Patterson F, et al. (2006). Nicotine metabolite ratio predicts efficacy of transdermal nicotine for smoking cessation. *Clin Pharmacol Ther*, 79(6):600-608.

Lesch OM, Dvorak A, Hertling I, et al. (2004). The Austrian multicentre study on smoking: subgroups of nicotine dependence and their craving. *Neuropsychobiology*, 50(1):78-88.

Lessov CN, Martin NG, Statham DJ, et al. (2004). Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med*, 34(5):865-879.

Levin ML, Goldstein H, Gerhardt PR (1950). Cancer and tobacco smoking; a preliminary report. *JAMA*, 143(4):336-338.

Levy DT, Bauer JE, Lee HR (2006). Simulation modeling and tobacco control: creating more robust public health policies. *Am J Public Health*, 96(3):494-498.

Lewis MJ, Wackowski O (2006). Dealing with an innovative industry: a look at flavored cigarettes promoted by mainstream brands. *Am J Public Health*, 96(2):244-251.

Ling PM, Landman A, Glantz SA (2002). It is time to abandon youth access tobacco programmes. *Tob Control*, 11(1):3-6.

Loken B, Howard-Pitney B (1988). Effectiveness of cigarette advertisements on women: an experimental study. *J Appl Psychol*, 73(3):378-382.

Loomis BR, Farrelly MC, Mann NH (2006). The association of retail promotions for cigarettes with the Master Settlement Agreement, tobacco control programmes and cigarette excise taxes**.** *Tob Control*, 15(6):458-463.

Lopez AD, Collishaw NE, Piha T (1994). A descriptive model of the cigarette epidemic in developed countries. *Tob Control*, 3:242-247.

Lovato C, Linn G, Stead LF, et al. (2003). *Impact of tobacco advertising and promotion on increasing adolescent smoking behaviours*. Cochrane Database Syst Rev, Issue 3. Art. No.: CD003439.

Ma JZ, Beuten J, Payne TJ, et al. (2005). Haplotype analysis indicates an association between the DOPA decarboxylase (DDC) gene and nicotine dependence. *Hum Mol Genet*, 14(12):1691-1698.

MacDonald G, Starr G, Schooley M, et al. (2001). *Introduction to program evaluation for comprehensive tobacco control programs*. Atlanta, GA, Centers for Disease Control and Prevention, 1-139. (*http://www.cdc.gov/tobacco/evaluation_manual/contents.htm*)

MacFadyen L, Hastings G, MacKintosh AM (2001). Cross sectional study of young people's awareness of and involvement with tobacco marketing. *Br Med J*, 322(7285):513-517.

MacKinnon DP, Lockwood CM, Hoffman JM, et al. (2002). A comparison of methods to test mediation and other intervening variable effects . *Psychol Methods*, 7(1):83-104.

Malouff J, Gabrilowitz D, Schutte N (1992). Readability of health warnings on alcohol and tobacco products. *Am J Public Health*, 82(3):464.

Marcus SE, Giovino GA, Pierce JP, et al. (1993). Measuring tobacco use among adolescents. *Public Health Rep*, 108(Suppl 1):20-24.

Marin G, Marin B (1989). A comparison of three interviewing approaches for studying sensitive topics with Hispanics. *Hisp J Behav Sci*, 11:330-340.

Marin G, Vanoss Marin B (1991). *Research with Hispanic Populations*. Newbury Park, CA, Sage Publications.

Mark MM, Henry GT, Julnes G (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco, Jossey Boss Inc.

Marketing Trends (1995). *Silk Cut Range Research*. Gallaher Ltd. Document No. 0572:1-4.

Marlow D, Crowne DP (1960). *The Approval Motive*. New York, John Wiley & Sons.

Mathieu JE, Taylor SR (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *J Organ Behav*, 27(8):1031-1056.

Matt GE, Wahlgren DR, Hovell MF, et al. (1999). Measuring environmental tobacco smoke exposure in infants and young children through urine cotinine and memory-based parental reports: empirical findings and discussion. *Tob Control*, 8(3):282-289.

Mayhew KP, Flay BR, Mott JA (2000). Stages in the development of adolescent smoking. *Drug Alcohol Depend*, 59(Suppl 1):S61-S81.

McDonald RP (1999). *Test theory: A unified treatment*. Mahwah, NJ, Lawrence Erlbaum Associates, Inc.

McEwen A, West R, Owen L, et al. (2005). General practitioners' views on and referral

to NHS smoking cessation services. *Public Health*, 119(4):262-268.

McGuinness T, Cowling K (1975). Advertising and aggregate demand for cigarettes. *Eur Econ Rev*, 6:311-328.

McKay RB, Breslow MJ, Sangster RL, et al. (1996). Translating Survey Questionnaires: Lessons Learned. *New Direct Eval*, (70):93-105.

McNeill A, Raw M, Whybrow J, et al. (2005). A national strategy for smoking cessation treatment in England. *Addiction*, 100(Suppl 2):1-11.

McNeill AD, Jarvis MJ, West R, et al. (1987). Saliva cotinine as an indicator of cigarette smoking in adolescents. *Br J Addict*, 82(12):1355-1360.

Mehl G, Stimpson GV, Riley L, et al. (2002). *Youth Tobacco - Rapid Assessment and Response Guide*. Geneva, World Health Organization.

Mekemson C, Glantz SA (2002). How the tobacco industry built its relationship with Hollywood. *Tob Control*, 11:I81-I91.

Merriman D (2001). *Understand, Measure and Combat Tobacco smuggling, Tool 7*. Smuggling. Washington, D.C., World Bank. *(http://www-lib/ doc-lib/TOB200703 Merriman2001.pdf)*

Merriman D, Yurekli A, Chaloupka FJ (2000). How big is the worldwide cigarette-smuggling problem? In: Jha P, Chaloupka FJ, eds., *Tobacco Control Policies in Developing Countries*. New York, Oxford University Press, 365-392.

Mirowsky J, Ross C (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2x2 index. *Soc Psychol Q*, 54(127):145.

Mochizuki-Kobayashi Y, Fishburn B, Baptiste J, et al.(2006). Use of cigarettes and other tobacco products among students aged 13-15 years—worldwide, 1999-2005. *Morb Mortal Wkly Rep*, 55(20):553-556.

Monninkhof E, van der Valk P, van der Palen J, et al. (2004). The effect of a minimal contact smoking cessation programme in out-patients with chronic obstructive pulmonary disease: a pre-post-test study. *Patient Educ Couns*, 52(3):231-236.

Mulcahy M, Evans DS, Hammond SK, et al. (2005). Secondhand smoke exposure and risk following the Irish smoking ban: an assessment of salivary cotinine concentrations in hotel workers and air nicotine levels in bars. *Tob Control*, 14(6):384-388.

Mullen PD, Carbonari JP, Tabak ER, et al. (1991). Improving disclosure of smoking by pregnant women. *Am J Obstet Gynecol*, 165(2):409-413.

Murray DM, Perry CL (1987). The measurement of substance use among adolescents: when is the 'bogus pipeline' method needed? *Addict Behav*, 12(3):225-233.

Murray RP, Bailey WC, Daniels K, et al. (1996). Safety of nicotine polacrilex gum used by 3,094 participants in the Lung Health Study. Lung Health Study Research Group. *Chest*, 109(2):438-445.

Muthen B, Asparouhov T (2006). Item response mixture modeling: application to tobacco dependence criteria. *Addict Behav*, 31(6):1050-1066.

Muthen LK, Muthen BO (2004). *Mplus User's Guide: Statistical Analysis with Latent Variables*. Los Angeles, Muthen & Muthen.

National Cancer Institute (2001). *Risks Associated with Smoking Cigarettes with Low Machine-Measured Yields of Tar and Nicotine*. Shopland DR, ed. Smoking and Tobacco Control Monograph No. 13. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub. No. 02-5047. *(http://cancercontrol.cancer.gov/tcrb/monographs/13/)*

National Cancer Institute (2002). *Making Health Communications Programs Work*. US Department of Health and Human Services, National Institutes of Health. Bethesda, MD, US Department of Health and Human Services, National Institute of Health.

National Cancer Institute (2005). *ASSIST: Shaping the Future of Tobacco Prevention and Control*. Marcus E, ed. Smoking and Tobacco Control Monograph No. 16. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub No. 05-5645. *(http://cancercontrol. cancer.gov/TCRB/monographs/16/index.html)*

National Cancer Institute (2006). *Evaluating ASSIST: A Blueprint for Understanding State-Level Tobacco Control*. Marcus E, ed. Smoking and Tobacco Control Monograph No. 17. Bethesda, MD, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, NIH Pub. No. 06-6085. *(http://cancercontrol.cancer.gov/tcrb/monographs/17/index.html)*

Navas-Acien A, Peruga A, Breysse P, et al. (2004). Secondhand tobacco smoke in public places in Latin America, 2002-2003. *JAMA*, 291(22):2741-2745.

Nebot M, Lopez MJ, Gorini G, et al. (2005). Environmental tobacco smoke exposure in public places of European cities. *Tob Control*, 14(1):60-63.

Nederhof A (1985). Methods of coping with social desirability: A review. *Eur J Soc Psychol*, 16:263-280.

Newcombe DA, Humeniuk RE, Ali R (2005). Validation of the World Health Organization Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): report of results from the Australian site. *Drug Alcohol Rev*, 24(3):217-226.

Niederdeppe J (2005). Assessing the validity of confirmed AD recall measures for public health communication campaign evaluation. *J Health Commun*, 10(7):635-650.

Niederdeppe J, Farrelly MC, Wenter D (2007). Media advocacy, tobacco control policy change and teen smoking in Florida. *Tob Control*, 16(1):47-52.

Noland MP, Kryscio RJ, Riggs RS, et al. (1988). Saliva cotinine and thiocyanate: chemical indicators of smokeless tobacco and cigarette use in adolescents. *J Behav Med*, 11(5):423-433.

Nunnally JC, Bernstein IH (1994). *Psychometric Theory*. New York, McGraw-Hill, Inc.

Nusser SM, Carriquiry AL, Dodd KW, et al. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *J Am Stat Assoc*, 91:1440-1449.

O'Connor RJ, Stitt JP, Kozlowski LT (2005). A digital image analysis system for identifying filter vent blocking on ultralight cigarettes. *Cancer Epidemiol Biomarkers Prev*, 14(2):533-537.

O'Connor RJ, Cummings KM, Giovino GA, et al. (2006a). How did UK cigarette makers reduce tar to 10 mg or less? *Br Med J*, 332(7536):302.

O'Connor RJ, Giovino GA, Fix BV, et al. (2006b). Smokers' reactions to reduced ignition propensity cigarettes. *Tob Control*, 15(1):45-49.

O'Connor RJ, Kozlowski LT, Borland R, et al. (2006c). Relationship between constituent labelling and reporting of tar yields among smokers in four countries. *J Public Health (Oxf)*, 28(4):324-329.

O'Connor RJ, Kozlowski LT, Hammond D, et al. (2007). Digital image analysis of cigarette filter staining to estimate smoke exposure. *Nicotine Tob Res*, 9(8):865-871.

O'Hegarty M, Pederson LL, Nelson DE, et al. (2006). Reactions of young adult smokers to warning labels on cigarette packages. *Am J Prev Med*, 30(6H):467-473.

O'Loughlin J, Paradis G, Renaud L, et al. (1998). One-year predictors of smoking initiation and of continued smoking among elementary schoolchildren in multiethnic, low-income, inner-city neighbourhoods. *Tob Control*, 7(3):268-275.

O'Loughlin J, DiFranza J, Tyndale RF, et al. (2003). Nicotine-dependence symptoms are associated with smoking frequency in adolescents. *Am J Prev Med*, 25(3):219-225.

O'Malley PM, Johnson LD, Bacham JG, et al. (2000). A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *J Drug Issues*, 30(1):35-54.

Ogawa M, Oyama T, Isse T, et al. (2006). Hemoglobin adducts as a marker of exposure to chemical substances, especially PRTR class I designated chemical substances. *J Occup Health*, 48(5):314-328.

Ogg CL (1964). Determination of particulate matter and alkaloids (as nicotine) in cigarette smoke. *J Assoc Official Agric Chems*, 47(2):356.

Ones D, Viswesvaran C, Reiss A (1996). Role of social desirability in personality testing for personnel selection: The red herring. *J Appl Psychol*, 81:660-679.

Oppenheim AN (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. London, Continuum.

Orzechowski W, Walker RC (2007). *The Tax Burden on Tobacco*. Arlington, Va: Orzechowski and Walker.

Ossip-Klein DJ, Giovino GA, Megahed N, et al. (1991). Effects of a smoker's hotline: results of a 10-county self-help trial. *J Consult Clin Psychol*, 59(2):325-332.

Owen L, McNeill A (2001). Saliva cotinine as indicator of cigarette smoking in pregnant women. *Addiction*, 96(7):1001-1006.

Owen L, Youdan B (2006). 22 years on: the impact and relevance of the UK No Smoking Day. *Tob Control*, 15(1):19-25.

Pan Y, de la Puente M (2005). *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation of How the Guideline was Developed*. Washington D.C., United States Bureau of the Census, Statistical Division.

Pankow JF, Tavakoli AD, Luo W, et al. (2003). Percent free base nicotine in the tobacco smoke particulate matter of selected commercial and reference cigarettes. *Chem Res Toxicol*, 16(8):1014-1018.

Parascandola M (2005). Lessons from the history of tobacco harm reduction: The National Cancer Institute's Smoking and Health Program and the "less hazardous cigarette." *Nicotine Tob Res*, 7(5):779-789.

Patrick DL, Cheadle A, Thompson DC, et al. (1994). The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health*, 84(7):1086-1093.

Patten CA, Martin JE, Calfas KJ, et al. (2001). Behavioral treatment for smokers with a history of alcoholism: predictors of successful outcome. *J Consult Clin Psychol*, 69(5):796-801.

Patton MQ (1997). *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks, CA, Sage Publications.

Paulhus D (1990). Measurement and Control of Response Bias. In: Robinson J, Shaver P, Wrightsman L, eds., *Measures of Personality and Social Psychological Attitudes.* San Diego, Academic Press, 17-59.

Payne TJ, Smith PO, McCracken LM, et al. (1994). Assessing nicotine dependence: a comparison of the Fagerstrom Tolerance Questionnaire (FTQ) with the Fagerstrom Test for Nicotine Dependence (FTND) in a clinical sample. *Addict Behav*, 19(3):307-317.

Pearson RW, Ross M, Dawes R (1992). Personal Recall and the Limits of Retrospective Questions in Surveys. In: Tanur JM, Social Science Research Council (U.S.) Committee on Cognition, eds., *Questions about Questions: Inquiries into the Cognitive Bases of Surveys.* New York, Russell Sage Foundation.

Pechacek TF, Babb S (2004). How acute and reversible are the cardiovascular risks of secondhand smoke? *Br Med J*, 328(7446):980-983.

Pechacek TF, Fox BH, Murray DM, et al. (1984). Review of techniques for measurement of smoking behavior. In: Matarazzo JD, Weiss SM, Herd JA, et al., eds., *Behavioral health: a handbook of health enhancement and disease prevention.* New York, John Wiley & Sons, 123-140.

Peck RM (2002). *Poverty: Equity Issues, tobacco, and the Poor*. Washington, DC, The International Bank for Reconstruction and Development, the World Bank.

Peele DM, Riddick MG, Edwards ME, et al. (2001). Formation of Tobacco-Specific Nitrosamines in Flue-Cured Tobacco. *Recent Adv Tob Sci*, 27:3-12.

Pekurinen M (1989). The demand for tobacco products in Finland. *Br J Addict*, 84(10):1183-1192.

Perez-Stable EJ, Herrera B, Jacob P, III, et al. (1998). Nicotine metabolism and intake in black and white smokers. *JAMA*, 280(2):152-156.

Pergadia ML, Heath AC, Martin NG, et al. (2006a). Genetic analyses of DSM-IV nicotine withdrawal in adult twins. *Psychol Med*, 36(7):963-972.

Pergadia ML, Heath AC, Agrawal A, et al. (2006b). The implications of simultaneous smoking initiation for inferences about the genetics of smoking behavior from twin data. *Behav Genet*, 36(4):567-576.

Perry CL, Eriksen M, Giovino G (1994). Tobacco use: a pediatric epidemic. *Tob Control*, 3(2):97-98.

Peters E, Romer D, Slovic P, et al. (2007). The impact and acceptability of Canadian style cigarette warning labels among US smokers and nonsmokers. *Nicotine Tob Res*, (4):473-481.

Peto R, Lopez A (2001). Future Worldwide Health Effects of Current Smoking Patterns. In: Koop CE, Pearson CE, Schwartz MR, eds., *Critical Issues in Global Health.* San Fransisco, Joosey-Bass, 154-161.

Peto R, Lopez AD, Boreham J, et al. (1992). Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet*, 339(8804): 1268-1278.

Phillips DH (2002). Smoking-related DNA and protein adducts in human tissues. *Carcinogenesis*, 23(12):1979-2004.

Piepe AP, Charlton P, Morey J, et al. (1986). Does sponsored sport lead to smoking among children? *Health Educ J*, 45(3):145-148.

Pierce JP, Gilpin EA (1995). A historical analysis of tobacco marketing and the uptake of smoking by youth in the United States: 1890-1977. *Health Psychol*, 14(6):500-508.

Pierce JP, Dwyer T, DiGiusto E, et al. (1987). Cotinine validation of self-reported smoking in commercially run community surveys. *J Chronic Dis*, 40(7):689-695.

Pierce JP, Fiore MC, Novotny TE, et al. (1989). Trends in cigarette smoking in the United States. Educational differences are increasing. *JAMA*, 261(1):56-60.

Pierce JP, Macaskill P, Hill D (1990). Long-term effectiveness of mass media led anti-smoking campaigns in Australia. *Am J Public Health*, 80(5):565-569.

Pierce JP, Lee L, Gilpin EA (1994). Smoking initiation by adolescent girls, 1944 through 1988. An association with targeted advertising. *JAMA*, 271(8):608-611.

Pierce JP, Choi WS, Gilpin EA, et al. (1996). Validation of susceptibility as a predictor of which adolescents take up smoking in the United States. *Health Psychol*, 15(5):355-361.

Pierce JP, Gilpin EA, Emery SL, et al. (1998a). Has the California tobacco control program reduced smoking? *JAMA*, 280(10):893-899.

Pierce JP, Farkas AJ, Gilpin EA (1998b). Beyond stages of change: the quitting continuum measures progress towards successful smoking cessation. *Addiction*, 93(2):277-286.

Pierce JP, Choi WS, Gilpin EA, et al. (1998c). Tobacco industry promotion of cigarettes and adolescent smoking. *JAMA*, 279(7):511-515.

Pierce JP, Gilpin EA, Choi WS (1999). Sharing the blame: smoking experimentation and future smoking-attributable mortality due to Joe Camel and Marlboro advertising and promotions. *Tob Control*, 8(1):37-44.

Pierce JP, Distefan JM, Kaplan RM, et al. (2005). The role of curiosity in smoking initiation. *Addict Behav*, 30(4):685-696.

Piper ME , McCarthy DE , Bolt DM ,et al. (2008). Assessing dimensions of nicotine dependence: an evaluation of the Nicotine Dependence Syndrome Scale (NDSS) and the Wisconsin Inventory of Smoking Dependence Motives (WISDM). *Nicotine Tob Res*, 10(6):1009-1020.

Piper ME, Piasecki TM, Federman EB, et al. (2004). A multiple motives approach to tobacco dependence: the Wisconsin Inventory of Smoking Dependence Motives (WISDM-68). *J Consult Clin Psychol*, 72(2):139-154.

Piper ME, McCarthy DE, Baker TB (2006). Assessing tobacco dependence: a guide to measure evaluation and selection. *Nicotine Tob Res*, 8(3):339-351.

Piper ME, Federman EB, McCarthy DE, et al. (2007). Efficacy of bupropion alone and in combination with nicotine gum. *Nicotine Tob Res*, 9(9):947-954.

Pirkle JL, Flegal KM, Bernert JT, et al. (1996). Exposure of the US population to environmental tobacco smoke: the Third National Health and Nutrition Examination Survey, 1988 to 1991. *JAMA*, 275(16): 1233-1240.

Pollay RW (2000). Targeting youth and concerned smokers: evidence from Canadian tobacco industry docu- ments. *Tob Control*, 9(2):136-147.

Pollay RW (2001). *The Role of Packaging seen through Industry Documents*. Montreal,

Quebec, Attorney General of Canada and Canadian Cancer Society, Supreme Court.

Pollay RW, Dewhirst T (2002). The dark side of marketing seemingly "Light" cigarettes: successful images and failed fact. *Tob Control*, 11 (Suppl. 1):i18-i31.

Pollay RW, Siddarth S, Siegel M, et al. (1996). The last straw cigarette advertising and realized market shares among youth and adults, 1979-1993. *J Marketing*, 60:1-16.

Polzin GM, Kosa-Maines RE, Ashley DL, et al. (2007). Analysis of volatile organic compounds in mainstream cigarette smoke. *Environ Sci Technol*, 41(4):1297-1302.

Pomerleau CS, Pomerleau OF, Flessland KA, et al. (1992). Relationship of Tridimensional Personality Questionnaire scores and smoking variables in female and male smokers. *J Subst Abuse*, 4(2):143-154.

Pomerleau CS, Carton SM, Lutzke ML, et al. (1994). Reliability of the Fagerstrom Tolerance Questionnaire and the Fagerstrom Test for Nicotine Dependence. *Addict Behav*, 19(1):33-39.

Portillo F, Antonanzas F (2002). Information disclosure and smoking risk perceptions. Potential short-term impact on Spanish students of the new European Union directive on tobacco products. *Eur J Public Health*, 12(4):295-301.

Post A, Gilljam H, Rosendahl I, et al. (2005). Validity of self reports in a cohort of Swedish adolescent smokers and smokeless tobacco (snus) users. *Tob Control*, 14(2):114-117.

Pound E, Coleman T, Adams C, et al. (2005). Targeting smokers in priority groups: the influence of government targets and policy statements. *Addiction*, 100 (Suppl 2):28-35.

Prochaska JO, DiClemente CC, Norcross JC (1992). In search of how people change. Applications to addictive behaviors. *Am Psychol*, 47(9):1102-1114.

Prochaska JO, Velicer WF, Rossi JS, et al. (1994). Stages of change and decisional balance for 12 problem behaviors. *Health Psychol*, 13(1):39-46.

Prokhorov AV, Winickoff JP, Ahluwalia JS, et al. (2006). Youth tobacco use: a global perspective for child health care clinicians. *Pediatrics*, 118(3):e890-e903.

Radke JM (1998). *Strategic Communications for Nonprofit Organizations: Seven Steps to Creating a Successful Plan*. New York, NY, Wiley.

Radzius A, Gallo JJ, Epstein DH, et al. (2003). A factor analysis of the Fagerstrom Test for Nicotine Dependence (FTND). *Nicotine Tob Res*, 5(2):255-40.

Radzius A, Gallo J, Gorelick D, et al. (2004). Nicotine dependence criteria of the DIS and DSM-III-R: a factor analysis. *Nicotine Tob Res*, 6(2):303-308.

Raitakari OT, Adams MR, McCredie RJ, et al. (1999). Arterial endothelial dysfunction related to passive smoking is potentially reversible in healthy young adults. *Ann Intern Med*, 130(7):578-581.

Ramsey LT, Pelletier A, Knight S (2004). Differences in smoking prevalence between the adult tobacco survey and the behavioral risk factor surveillance system. *Prev Chronic Dis* , 1(4):A22.

Randolph W, Viswanath K (2004). Lessons learned from public health mass media campaigns: marketing health in a crowded media world. *Annu Rev Public Health*, 25:419-437.

Ranson K, Jha P, Chaloupka FK, et al. (2000).The effectiveness and cost-effectiveness of price increases and other tobacco-control policies. In: Jha P, Chaloupka FK, eds., *Tobacco control in developing countries.* Oxford, Oxford University Press; 427-447.

Raw M, McNeill A, West R (1998). Smoking cessation guidelines for health professionals. A guide to effective smoking cessation interventions for the health care system. Health Education Authority. *Thorax*, 53 (Suppl 5 Pt 1):S1-19.

Raw M, Anderson P, Batra A, et al. (2002). WHO Europe evidence based recommendations on the treatment of tobacco dependence. *Tob Control*, 11(1):44-46.

Raw M, McNeill A, Coleman T (2005). Lessons from the English smoking treatment services. *Addiction*, 100 Suppl 2:84-91.

Razavi D, Vandecasteele H, Primo C, et al. (1999). Maintaining abstinence from cigarette smoking: effectiveness of group counselling and factors predicting outcome. *Eur J Cancer*, 35(8):1238-1247.

Rebagliato M, Bolumar F, Florey CD, et al. (1998). Variations in cotinine levels in smokers during and after pregnancy. *Am J Obstet Gynecol*, 178(3):568-571.

Redmond WH (1999). Effects of sales promotion on smoking among U.S. ninth graders. *Prev Med*, 28(3):243-250.

Reid JR (1993). A history of mentholated cigarettes. "This Spud's For You.". *Recent Advances in Tobacco Science*, 19:71-84.

Reinert DF, Allen JP (2002). The Alcohol Use Disorders Identification Test (AUDIT): a review of recent research. *Alcohol Clin Exp Res*, 26(2):272-279.

Repace J (2004). Respirable particles and carcinogens in the air of Delaware hospitality venues before and after a smoking ban. *J Occup Environ Med*, 46(9):887-905.

Richards JW, Fischer P, Conner FG (1989). The warnings on cigarette packages are ineffective. *JAMA*, 261(1):45.

Richter P, Spierto FW (2003). Surveillance of smokeless tobacco nicotine, pH, moisture, and unprotonated nicotine content. *Nicotine Tob Res*, 5(6):885-889.

Rijo MJ. (2005). *Price elasticity estimates for tobacco and other addictive goods in India*. Working Paper Series No. WP-2005-

003. Mumbai, India, Indira Gandhi Institute of Development Research.

Roberts DL (1988). Natural tobacco flavor. *Recent Advances in Tobacco Science*, 49-81.

Robinson TN, Killen JD (1997). Do cigarette warning labels reduce smoking? Paradoxical effects among adolescents. *Arch Pediatr Adolesc Med*, 151(3):267-272.

Rogers R (1975). A protection motivation theory of fear appeals and attitude change. *J Psychol*, 91:93-114.

Rogers R, Prentice-Dunn S (1997). Determinants of health behavior: Personal and social. Protection motivation theory. In: Gochman DS, ed., *Handbook of Health Behavior Research.* New York, Plenum: 113-132.

Rootman I, Flay BR, Phil D. (1995). *A Study on Youth Smoking: Plain Packaging, Health Warnings, Event Marketing and Price Reductions. Key Findings and Figures*. Toronto, University of Toronto.

Rootman I, Goodstadt M, Hyndman B, et al. (2001). *Evaluation in health promotion: principles and perspectives*. World Health Organization Regional Office for Europe. Copen- hagen, World Health Organization.

Ross CE, Mirowsky J (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *J Health Soc Behav*, 25(2):189-197.

Ross H, Chaloupka FJ (2006). Economic policies for tobacco control in developing countries. *Salud Publica Mex*, 48 (Suppl 1):S113-S120.

Ross M (1989). Relation of Implicit Theories to the Construction of Personal Histories. *Psychological Review*, 96(2):341-357.

Rossi PH, Freeman HE, Lipsey MW (2003). *Evaluation: a systematic approach*. Thousand Oaks, CA, Sage Publications.

Roswell Park Cancer Institute (2007). Tobacco Free Air. *(http://www.tobaccofreeair.org)*.

Roth D.H., Roth AB, Liu X (2005). Health risks of smoking compared to Swedish snus. *Inhal Toxicol*, 17(13):741-748.

Rothman KJ, Greenland S (1998). *Modern epidemiology*. Philadelphia, PA, Lippincott-Raven.

Rothmans (UK) Marketing Services. (1998). *Young Adult Smokers: Smoking Behaviour and Lifestyles 1994-1997*. Document No. 0626:1-26.

Rowena JH, Gale F, Capehart TC, et al. (2000). The supply-side effects of tobacco-control policies. In: Jha P, Chaloupka FK, eds., *Tobacco control in developing countries.* Oxford, Oxford University Press.

Royal College of Physicians. (2000). Nicotine Addiction in Britain. A Report of the Tobacco Advisory Group of the RCP. *(http://www.rcplondon.ac.uk/pubs/ books/ nicotine)*

Ruel E, Mani N, Molnar E, et al. (2001). *Retailer Practices and Youth Tobacco Access: What Communities can do to Minimize the Influence of Tobacco Industry, ImpacTeen.(http://www.impacteen.org/generalarea_PDFs/NCTOH2001_reul.pdf)*

Rumpf HJ, Hapke U, Meyer C, et al. (2002). Screening for alcohol use disorders and at-risk drinking in the general population: psychometric performance of three questionnaires. *Alcohol Alcohol*, 37(3): 261-268.

Saccone SF, Hinrichs AL, Saccone NL, et al. (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*, 16(1):36-49.

Saffer H, Chaloupka F (2000). The effect of tobacco advertising bans on tobacco consumption. *J Health Econs*, 19(6):1117-1137.

Sargent JD, Dalton M, Beach M (2000). Exposure to cigarette promotions and smoking uptake in adolescents: evidence of a dose-response relation. *Tob Control*, 9(2):163-168.

Saris W (2003a). Multitrait-multimethod studies. In: Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Survey Methods*. Hoboken, NJ, Wiley & Sons: 265-274.

Saris W (2003b). Response Functioin Equality. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Survey Research.* Hobokin, NJ, Wiley & Sons: 275-289.

Sarntisart I. (2003). *An economic analysis of tobacco control in Thailand*. World Bank. Health, Nutrition and Population Discussion Paper. Economics of Tobacco Control Paper No. 15. (*http://www1.worldbank.org/tobacco/pdf/Sarntisart-An%20Economic_Thailand_new.pdf*)

Sayginsoy O, Yurekli A, De Beyer J. (2002). *Cigarette Demand, Taxation and the Poor: A case stuy of Bulgaria*. World Bank. Health, Nutrition and Population Discussion Paper. Economics of Tobacco Control Paper No. 4. *(http://www1.worldbank.org/tobacco/pdf/Sayginsoy-%20Cigarette%20Bulgaria-whole.pdf*)

Schmidt CW (2006). Signs of the times: biomarkers in perspective. *Environ Health Perspect*, 114(12):A700-A705.

Schooler C, Basil MD, Altman DG (1996). Alcohol and cigarette advertising on billboards: Targeting with social clues. *Health Commun*, 8:109-129.

Schultz DE, Kitchen J (2000). *Communicating Globally: An Integrated Marketing Approach.* Maidenhead, McGraw-Hill.

Sciamanna CN, Hoch JS, Duke GC, et al. (2000). Comparison of five measures of motivation to quit smoking among a sample of hospitalized smokers. *J Gen Intern Med*, 15(1):16-23.

Scollo M, Lal A, Hyland A, et al. (2003). Review of the quality of studies on the economic effects of smoke-free policies on the hospitality industry. *Tob Control*, 12(1):13-20.

Sepe E, Ling PM, Glantz SA (2002). Smooth moves: bar and nightclub tobacco promotions that target young adults. *Am J Public Health*, 92(3):414-419.

Severson HH, Ary DV (1983). Sampling bias due to consent procedures with adolescents. *Addict Behav*, 8(4):433-437.

Shadish WR, Cook TD, Campbell DT (2002). *Experimental and quasi-experimental Designs for Generalized Causal Inference*. Boston, MI, Houghton Mifflin.

Shafey O, Cokkinides V, Cavalcante TM, et al. (2002). Case studies in international tobacco surveillance: cigarette smuggling in Brazil. *Tob Control*, 11(3):215-219.

Shafey O, Dolwick S, Guindon GE (2003). Tobacco control country profiles. Atlanta, American Cancer Society, World Health Organization and International Union Against Cancer. *(http://www.who.int/tobacco/global_data/country_profiles/en/index.html)*

Shapiro S, MacInnis DJ, Heckler SE (1997). The effects of incidental and exposure on the formation of consideration sets. *J Consumer Res*, 24:94-104.

Shiffman S, Sayette MA (2005). Validation of the nicotine dependence syndrome scale (NDSS): a criterion-group design contrasting chippers and regular smokers. *Drug Alcohol Depend*, 79(1):45-52.

Shiffman S, Pillitteri JL, Burton SL, et al. (2001). Smokers' beliefs about "Light" and "Ultra Light" cigarettes. *Tob Control*, 10 Suppl. 1:i17-i23.

Shiffman S, Gwaltney CJ, Balabanis MH, et al. (2002). Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. *J Abnorm Psychol*, 111(4):531-545.

Shiffman S, Hughes JR, Pillitteri JL, et al. (2003). Persistent use of nicotine replacement therapy: an analysis of actual purchase patterns in a population based sample. *Tob Control*, 12(3):310-316.

Shiffman S, Waters A, Hickcox M (2004). The nicotine dependence syndrome scale: a multidimensional measure of nicotine dependence. *Nicotine Tob Res*, 6(2):327-348.

Shillington AM, Clapp JD (2000). Self-report stability of adolescent substance use: are there differences for gender, ethnicity and age? *Drug Alcohol Depend*, 60(1):19-27.

Shopland DR, Anderson CM, Burns DM, et al. (2004).Disparities in smoke-free workplace policies among food service workers. *J Occup Environ Med*, 46(4):347-356.

Siahpush M, McNeill A, Borland R, et al. (2006). Socioeconomic variations in nicotine dependence, self-efficacy, and intention to quit across four countries: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15 (Suppl 3):iii71-iii75.

Siegel M, Nelson DE, Peddicord JP, et al. (1996). The extent of cigarette brand and company switching: results from the Adult Use-of-Tobacco Survey. *Am J Prev Med*, 12(1):14-16.

Sillet RW, Wilson MB, Malcom RE, et al. (1978). Deception among smokers. *Br Med J*, 2:1185-1186.

Skeer M, Land ML, Cheng DM, et al. (2004). Smoking in Boston bars before and after a 100% smoke-free regulation: an assessment of early compliance. *J Public Health Manag Pract*, 10(6):501-507.

Slade J (1997). The pack as advertisement. *Tob Control*, 6(3):169-170.

Slade J (2001). Marketing Policies. In: Rabin RL, Sugarman SD, eds., *Regulating tobacco.* Oxford, Oxford University Press: 72-110.

Slater SJ, Chaloupka FJ, Wakefield M, et al. (2007). The impact of retail cigarette marketing practices on youth smoking uptake. *Arch Pediatr Adolesc Med*, 161(5):440-445.

Sledjeski EM, Dierker LC, Costello D, et al. (2007). Predictive validity of four nicotine dependence measures in a college sample. *Drug Alcohol Depend*, 87(1):10-19.

Slovic P (1998). Do adolescent smokers know the risks? *Duke Law J*, 47(6):1133-1141.

Sly DF, Hopkins RS, Trapido E, et al. (2001a). Influence of a counteradvertising media campaign on initiation of smoking: the Florida "truth" campaign. *Am J Public Health*, 91(2):233-238.

Sly DF, Heald GR, Ray S (2001b). The Florida "truth" anti-tobacco media evaluation: design, first year results, and implications for planning future state media evaluations. *Tob Control*, 10(1):9-15.

Smee C, Parsonage M, Anderson R, et al. (1992). *Effect of tobacco advertising on tobacco consumption: A discussion document reviewing the evidence*. London, Economics & Operational Research Division, Department of Health.

Smith A (1776). *An inquiry into the nature and causes of the wealth of nations 43*. Chicago, University of Chicago Press.

Smith P (2004b). Acquiescent response bias as an aspect of cultural communication style. *J Cross Cult Psychol*, 35:50-61.

Smith TW (2004a). Developing and evaluating cross-national survey instruments. In: Pressler S, Rothgeb JM, Couper MP, et al., eds., *Methods for Testing and Evaluating Survey Questionnaires.* Hoboken, NJ, Wiley & Sons: 431-452.

Snyder LB (2001). How Effective Are Mediated Health Campaigns? In: Rice RE, Atkin CK, eds., *Public communication campaigns.* Thousand Oaks, CA, Sage Publications.

Song S, Ashley DL (1999). Supercritical fluid extraction and gas chromatography/mass spectrometry for the analysis of tobacco-specific nitrosamines in cigarettes. *Anal Chem*, 71(7):1303-1308.

Spencer SJ, Zanna MP, Fong GT (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *J Pers Soc Psychol*, 89(6):845-851.

Spradley JP (1979). *Ethnographic Interview.* New York, Holt, Rinehart and Winston.

SRNT Subcommittee on Biochemical Verification (2002). Biochemical Verification of Tobacco Use and Cessation. *Nicotine Tob Res*, 4:149-159.

Stanfill SB, Ashley DL (2000). Quantitation of flavor-related alkenylbenzenes in tobacco smoke particulate by selected ion monitoring gas chromatography-mass spectrometry. *J Agric Food Chem*, 48(4):1298-1306.

Starr G, Rogers T, Schooley M, et al. (2005). *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs.* Atlanta, GA, Centers for Disease Control and Prevention. *(http://www.cdc.gov/TOBACCO/-tobacco_ control_programs/surveillance_evaluation/key_outcome/index.htm)*

Stead LF, Lancaster T (2000). A systematic review of interventions for preventing tobacco sales to minors. *Tob Control*, 9(2):169-176.

Stein LA, Colby SM, O'Leary TA, et al. (2002). Response distortion in adolescents who smoke: a pilot study. *J Drug Educ*, 32(4):271-286.

Stephenson MT, Hoyle RH, Palmgreen P, et al. (2003). Brief measures of sensation seeking for screening and large-scale surveys. *Drug Alcohol Depend*, 72(3):279-286.

STEPS. (2007). The Stepwise Approach to Surveillance. (*http://www.who.int/chp/steps/en)*.

Stewart DW, Shamdasani PN (1998). Focus group research: Exploration and discovery. In: Bickman L, Rog DJ, eds., *Handbook of Applied Social Research Methods.* Thousand Oaks, CA, Sage Publications: 505-526.

Storr CL, Reboussin BA, Anthony JC (2005). The Fagerstrom test for nicotine dependence: a comparison of standard scoring and latent class analysis approaches. *Drug Alcohol Depend*, 80(2):241-250.

Strahan EJ, White K, Fong GT, et al. (2002). Enhancing the effectiveness of tobacco package warning labels: a social psychological perspective. *Tob Control*, 11(3):183-190.

Straif K, Baan R, Grosse Y, et al. (2005). Carcinogenicity of polycyclic aromatic hydrocarbons. *Lancet Oncol*, 6(12):931-932.

Strasser AA, O'Connor RJ, Mooney ME, et al. (2006). Digital image analysis of cigarette filter stains as an indicator of compensatory smoking. *Cancer Epidemiol Biomarkers Prev*, 15(12):2565-2569.

Strong DR, Kahler CW, Ramsey SE, et al. (2003). Finding order in the DSM-IV nicotine dependence syndrome: a Rasch analysis. *Drug Alcohol Depend*, 72(2):151-162.

Sunley E, Yurekli A, Chaloupka F (2000). The design, administration and potential revenue of tobaccob excises. In: Jha P, Chaloupka F, eds., *Tobacco in Developing Countries.* Oxford, Oxford University Press.

Sutherland M, Galloway J (1981). Role of advertising: Persuasion or agenda setting? *J Advert Res*, 21(5):25-29.

Swets JA, Dawes RM, Monahan J (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, (1):1-26.

Tauras JA (2004). Public policy and smoking cessation among young adults in the United States. *Health Policy*, 68(3):321-332.

Tauras JA, Chaloupka F (2001). Determinants of smoking cessation: an analysis of young adult men and women. In: Grossman M, Hsieh CR, Northampton MA, eds., *The Economic Analysis of Substance Use and Abuse: The Experience of Developed countries and Lessons for Developing countries.* Edward Elgar Publishing.

Tauras JA, Johnson LD, O'Malley PM. (2001). *Effects of Price and Access Laws on Teenage smoking Initiation: A National Longitudinal Analysis.* Cambridge, National Bureau of Economic Research.

Tauras JA, Peck RM, Chaloupka FJ (2006). The role of retail prices and promotions in determining cigarette brand market shares. *Rev Ind Organ*, 28(3):253-284.

Taylor AL, Bettcher DW (2000). WHO Framework Convention on Tobacco Control: a global "good" for public health. *Bull World Health Organ*, 78(7):920-929.

Thomas R, Perera R (2006). School-based programmes for preventing smoking. *Cochrane Database Syst Rev*, 3:CD001293.

Thompson ME, Fong GT, Hammond D, et al. (2006). Methods of the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15 (Suppl 3):iii12-iii18.

Thrasher JF, Bentley ME (2006). The meanings and context of smoking among Mexican university students. *Public Health Rep*, 121(5):578-585.

Thrasher JF, Jackson C (2006). Mistrusting companies, mistrusting the tobacco industry: clarifying the context of tobacco prevention efforts that focus on the tobacco industry. *J Health Soc Behav*, 47(4):406-422.

Thrasher JF, Cummings KM, Warnecke R, et al. (2006a). *Clarifying the Concept of "Denormalization" in Tobacco Prevention Efforts.* 3rd World Conference on Tobacco or Health.

Thrasher JF, Niederdeppe JD, Jackson C, et al. (2006b). Using anti-tobacco industry messages to prevent smoking among high-risk adolescents. *Health Educ Res*, 21(3):325-337.

Thrasher JF, Rousu MC, Ocampo-Anaya R, et al. (2007). Estimating the impact of different package warning label policies: the auction method. *Addict Behav*, 32(12):2916-2925.

Thun MJ, Jemal A (2006). How much of the decrease in cancer death rates in the United States is attributable to reductions in tobacco smoking? *Tob Control*, 15(5): 345-347.

Tobacco Institute. (1991). Tobacco Tax Survey 1991. Available on-line at: *http://to-*

baccodocuments.org/nysa_ti_m2/ TI38980 378. htlm

Tomar SL, Giovino GA, Eriksen MP (1995). Smokeless tobacco brand preference and brand switching among US adolescents and young adults. *Tob Control*, 4(1):67-72.

Townsend J, Roderick P, Cooper J (1994). Cigarette smoking by socioeconomic group, sex, and age: effects of price, income, and health publicity. *Br Med J*, 309(6959):923-927.

Travers MK, Cummings KM, Hyland A, et al. (2004). Indoor Air Quality in Hospitality Venues Before and After Implementation of a Clean Indoor Air Law. *Morb Mortal Wkly Rep*, 1038.

Triandis HC (1995). *Individualism and Collectivism*. Boulder, CO, Westview Press.

Triandis HC, Gelfand MJ (1998). Converging measurement of horizontal and vertical individualism and collectivism. *J Pers Soc Psychol*, 74(1):118-128.

Trochim WMK (1984). *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Newbury Park, CA, Sage Publications.

Trochim WMK, Cappelleri JC, Reichardt CS (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: II. When an interaction effect is present. *Eval Rev*, 15:571-604.

Trosclair A, Caraballo R, Malarcher A, et al. (2005). Cigarette smoking among adults - United States, 2003 (Reprinted from MMWR, vol 54, pg 509-513, 2005). *JAMA*, 294(2):172-173.

Transdisciplanary Tobacco Use Research Centre (TTURC)TURC Tobacco Dependence Phenotype Workgroup (2007). Time to first cigarette in the morning as an index of ability to quit smoking: Implications for nicotine dependence. *Nicotine Tob Res*, 9(S4):S555-S570.

Turner CF, Lessler JT, Devore JW (1992). Effects of mode of administration and wording on reporting of drug use. In: Turner CF, Lessler JT, Gfroerer J, eds., *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD, U.S. Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.

Tverdal A, Bjartveit K (2006). Health consequences of reduced daily cigarette consumption. *Tob Control*, 15(6):472-480.

Tyas SL, Pederson LL (1998). Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tob Control*, 7(4):409-420.

UK Laboratory of the Government Chemist (2000). Comparison of benzo[a]pyrene yields from USA and UK cigarette blends. In: *Cigarette Yields Using Intense Smoking Protocols* London,UK, LGC Report FN40/M26/00.

U.S. Department of Health, Education, and Welfare (1964). *Smoking and health report of the advisory committee to the Surgeon General of the Public Health Service*. Washington, *D.C.,* U.S. Dept. of Health, Education, and Welfare, Public Health Service.

U.S.Department of Health and Human Services. (1986). The Health Consequences of Involuntary Smoking: A Report of the Surgeon General, Rockville, MD, U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (1988). *The Health Consequences of Smoking: Nicotine Addiction: A Report of the Surgeon General*. Rockville, MD, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (1989). *Reducing the Health Consequences of Smoking: 25 Years of Progress: A Report of the Surgeon General: 1989 Executive Summary*. Rockville, MD, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (1990). *The Health Benefits of Smoking Cessation : A Report of the Surgeon General.* Rockville, MD, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (1994). *Preventing Tobacco Use among Young People: A Report of the Surgeon General*. Atlanta, GA., U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (1998). *Tobacco Use Among U.S. Racial/Ethnic Minority Groups: African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, Hispanics : A Report of the Surgeon General.* Atlanta, GA, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (2001). *Women and Smoking: A Report of the Surgeon General*. Rockville, MD, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (2004). *The Health Consequences of Smoking: A Report of the Surgeon General*. Rockville, MD, U.S. Dept. of Health and Human Services, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. (2006). *The Health Consequences of Involuntary Exposure to Tobacco Smoke : A Report of the Surgeon General.* Rockville, MD, U.S. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S.Department of Health and Human Services. (2008). *Treating Tobacco Use and Dependence: 2008 Update. U.S.* Department of Health and Human Services, Public Health Service; 2008 May. 257 p.

United Nations Statistical Division (2003). *Industrial Commodity Statistics Yearbook*. New York, NY, United Nations, Dlept of Economic and Policy Analysis.

Van de Vijver FJR (2003). Bias and substantive analyses. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Survey Methods.* Hoboken, NJ, Wiley & Sons: 207-233.

Van de Vijver FJR (2004). Bias and Equivalence: Cross-Cultural Perspectives. In: Harkness JA, Van de Vijver FJR, Mohler PP, eds., *Cross-Cultural Surveys Methods.* Hoboken, NH, Wiley & Sons

Van de Vijver F, Hambleton RK (1996). Translating Tests: Some pratical Guidelines. *European Psychol*, 1(2):89-99.

Van de Vijver F, Leung K (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA, Sage Publications.

Van der Zouwen J, Smit JH (2004). Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach. In: Pressler S, Rothgeb JM, Couper MP, et al., eds., *Methods for Testing and Evaluating Survey Questionnaires.* Hoboken, NJ, Wiley & Sons: 109-130.

Vartiainen E, Seppala T, Lillsunde P, et al. (2002). Validation of self reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Commu Health*, 56(3):167-170.

Velicer WF, DiClemente CC, Prochaska JO, et al. (1985). Decisional balance measure for assessing and predicting smoking status. *J Pers Soc Psychol*, 48(5):1279-1289.

Velicer WF, DiClemente CC, Rossi JS, et al. (1990). Relapse situations and self-efficacy: an integrative model. *Addict Behav*, 15(3):271-283.

Velicer WF, Prochaska JO, Rossi JS, et al. (1992). Assessing outcome in smoking cessation studies. *Psychol Bull*, 111(1):23-41.

Vijayalaxmi, Evans HJ (1982). In vivo and in vitro effects of cigarette smoke on chromosomal damage and sister-chromatid exchange in human peripheral blood lymphocytes. *Mutat Res*, 92(1-2):321-332.

Volk HE, Scherrer JF, Bucholz KK, et al. (2007). Evidence for specificity of transmission of alcohol and nicotine dependence in an offspring of twins design. *Drug Alcohol Depend*, 87(2-3):225-232.

Voorhees CC, Yanak LR, Stillman FA, et al. (1998). Reducing cigarette sales to minors in an urban setting: Issues and opportunities for merchant intervention. *Am J Prev Med*, 14:138-142.

Wakefield MA, Chaloupka FJ (1998). Improving the measurement and use of tobacco control "inputs". *Tob Control*, 7(4):333-335.

Wakefield M, Carrangis J, Wilson D, et al. (1992). Illegal cigarette sales to children in South Australia. *Tob Control*, (1):-114.

Wakefield M, Morley C, Horan JK, et al. (2002). The cigarette pack as image: new evidence from tobacco industry documents. *Tob Control*, 11 (Suppl. 1):I73-I80.

Wakefield M, Trotter L, Cameron M, et al. (2003a). Association between exposure to workplace secondhand smoke and reported respiratory and sensory symptoms: cross-sectional study. *J Occup Environ Med*, 45(6):622-627.

Wakefield M, Freeman J, Donovan R (2003b). Recall and response of smokers and recent quitters to the Australian National Tobacco Campaign. *Tob Control*, 12 (Suppl 2):ii15-ii22.

Wakefield M, McLeod K, Perry CL (2006). "Stay away from them until you're old enough to make a decision": tobacco company testimony about youth smoking initiation. *Tob Control*, 15 (Suppl 4):iv44-iv53.

Wallack L, Dorfman L (2001). Putting policy into health communication: The roles of media advocacy. In: Rice RE, Atkin CK, eds., *Public Communication Campaigns.* Thousand Oaks, CA, Sage Publications: 389-401.

Wallack L, Dorfman L, Jernigan D, et al. (1993). *Media Advocacy and Public Health: Power for Prevention.* Newbury Park, CA, Sage Publications. (http://www.amazon.com/gp-/reader/0803942893/ref=sib_dp_pt#reader-link)

Wallack L, Woodruff K, Dorfman L, et al. (1999). *News for a change: An Advocates' Guide to Working with the Media*. Thousand Oaks, CA, Sage Publications. *(http://www.amazon.com/gp/reader/0761919244/ref=sib_dp_pt#reader-link)*

Warnecke RB, Johnson TP, Chavez N, et al. (1997). Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol*, 7(5):334-342.

Warner KE (1977). The effects of the anti-smoking campaign on cigarette consumption. *Am J Public Health*, 67(7):645-650.

Warner KE (1978). Possible increases in the underreporting of cigarette consumption. *J Am Stat Assn*, 73:314-318.

Warner KE (2006). Tobacco Policy Research: Insights and Contributions to Public Health Policy. In: Warner KE, ed., *Tobacco Control Policy.* San Francisco, Jossey-Bass: 3-86.

Warren CW, Riley L, Asma S, et al. (2000). Tobacco use by youth: a surveillance report from the Global Youth Tobacco Survey project. *Bull World Health Organ*, 78(7):868-876.

Warren CW, Jones NR, Eriksen MP, et al. (2006). Patterns of global tobacco use in young people and implications for future chronic disease burden in adults. *Lancet*, 367(9512):749-753.

Watson C, McCraw J, Polzin G, et al. (2004a). Development of a method to assess cigarette smoke intake. *Environ Sci Technol*, 38(1):248-253.

Watson CH, Trommel JS, Ashley DL (2004b). Solid-phase microextraction-based approach to determine free-base nicotine in trapped mainstream cigarette smoke total particulate matter. *J Agric Food Chem*, 52(24):7240-7245.

Watson D (1992). Correcting for acquiescent response bias in the absence of a balanced scale. *Sociol Methods Res*, 21:52-88.

Wayne GF, Connolly GN (2002). How cigarette design can affect youth initiation into smoking: Camel cigarettes 1983-93. *Tob Control*, 11 (Suppl 1):I32-I39.

Weber MD, Bagwell DA, Fielding JE, et al. (2003). Long term compliance with California's Smoke-Free Workplace Law among bars and restaurants in Los Angeles County. *Tob Control*, 12(3):269-273.

Wegner DM (1994). Ironic processes of mental control. *Psychol Rev*, 101(1):34-52.

Weidmer B, Agency for Healthcare Research and Quality. (2006). *Guidelines for translating CAHPS surveys. U.S*. Dept. of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Weinstein ND (1999). Accuracy of smokers' risk perception. *Nicotine Tob Res*, 1 (Suppl 1):S123-S130.

Werner L, Campbell DT (1970). Translating, working through interpreters and the problem of decentering. In: Naroll R, Cohen R, eds., *Handbook of Methods in Cultural Anthropology.* Garden City, NY, Natural History Press: 398-420.

West R (2005a). Time for a change: putting the Transtheoretical (Stages of Change) Model to rest. *Addiction*, 100(8):1036-1039.

West R. (2005b). *Assessing Smoking Cessation Performance in NHS Stoop Smoking Services*. The Russell Standard (Clinical). *(http://www.aspsilverbackwebsites.co.uk/RobertWest/resources/russellstandardclinicalversion2.pdf)*

West R. (2006). *Feasibility of a national longitudinal study ("The Smoking Toolkit Study") to monitor smoking cessation and attempts at harm reduction in the UK*, 1-35. *(http://www.aspsilverbackwebsites.co.uk/smokinginengland/Ref/stp001.pdf)*

West R, Hardy A (2006). *Theory of addiction*. Oxford, Blackwell Pub./Addiction Press.

West R, Sohal T (2006). "Catastrophic" pathways to smoking cessation: findings from national survey. *Br Med J*, 332(7539):458-460.

West R, McNeill A, Raw M (2000). Smoking cessation guidelines for health professionals: an update. Health Education Authority. *Thorax*, 55(12):987-999.

West R, McEwen A, Bolling K, et al. (2001). Smoking cessation and smoking patterns in the general population: a 1-year follow-up. *Addiction*, 96(6):891-902.

West R, DiMarino ME, Gitchell J, et al. (2005). Impact of UK policy initiatives on use of medicines to aid smoking cessation. *Tob Control*, 14(3):166-171.

West R, Zatonski W, Przewozniak K, et al. (2007). Can we trust national smoking prevalence figures? Discrepancies between biochemically assessed and self-reported smoking rates in three countries. *Cancer Epidemiol Biomarkers Prev*, 16(4):820-822.

Westman EC, Behm FM, Simel DL, et al. (1997). Smoking behavior on the first day of a quit attempt predicts long-term abstinence. *Arch Intern Med*, 157(3):335-340.

Wewers ME, Dhatt RK, Moeschberger ML, et al. (1995). Misclassification of smoking status among Southeast Asian adult immigrants. *Am J Respir Crit Care Med*, 152(6 Pt 1):1917-1921.

White V, Hayman J. (2006). *Smoking behaviours of Australian secondary students in 2005*. Victoria, Australia, The Cancer Council Victoria.

White V, Hill D, Siahpush M, et al. (2003). How has the prevalence of cigarette smoking changed among Australian adults? Trends in smoking prevalence between 1980 and 2001. *Tob Control*, 12(Suppl 2):ii67-ii74.

White VM, Hill DJ, Effendi Y (2004). How does active parental consent influence the findings of drug-use surveys in schools? *Eval Rev*, 28(3):246-260.

WHO (1992). World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva, World Health Organization.

WHO (1993). World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. Geneva, World Health Organization.

WHO (1997). World Health Organization. *Tobacco or Health: A Global Status Report*. Geneva, World Health Organization.

WHO (1998a).World Health Organization. *Guidelines for Controlling and Monitoring the Tobacco Epidemic*. Geneva, World Health Organization.

WHO (1998b). World Health Organization. *Health Promotion Evaluation: Recommendations to Pollicy-Makers*. Report of the WHO European Working Group on Health Promotion Evaluation. 1-8.

WHO (2000). World Health Organization. *Advancing Knowledge on Regulating Tobacco Products*. (*http://www who int/tobacco/fctc/cop/en)*.

WHO (2003). World Health Organization. *WHO Framework Convention on Tobacco Control*. WHO Geneva.

WHO (2004). World Health Organization. *Basic Principles of Media Advocacy*. WHO Geneva.

WHO (2007a). World Health Organization. *Global school-based Student Health Survey country fact sheet,* 2007. (*http://www who int/chp/gshs/ facts-heets/en/index html)*.

WHO (2007b). World Health Organization. *The European Regional Office: Declaration on young people and alcohol, 2001. (http://who.euro.who. int/AboutWHO/Policy/20030204_1)*

WHO ASSIST Working Group (2002). The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction*, 97(9):1183-1194.

WHO Study Group on Tobacco Product Regulation. (2004). Recommendation: Guiding Principles for the Development of Tobacco Product Research and Testing Capacity and Proposed Protocols for the Initiation of Tobacco Product Testing. 1-42 (*http://www who int/tobacco/global_interaction/tobreg/goa_2003_principles/en/index html)*.

Whooley MA, Avins AL, Miranda J, et al. (1997). Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med*, 12(7):439-445.

Wiehe SE, Garrison MM, Christakis DA, et al. (2005). A systematic review of school-based smoking prevention trials with long-term follow-up. *J Adolesc Health*, 36(3):162-169.

Willemsen MC (2005). The new EU cigarette health warnings benefit smokers who want to quit the habit: results from the Dutch Continuous Survey of Smoking Habits. *Eur J Public Health*, 15(4):389-392.

Willemsen MC, Simons C, Zeeman G (2002). Impact of the new EU health warnings on the Dutch quit line. *Tob Control*, 11(4):381-382.

Williams WS, Montgomery K, Pasnik S (1997). *Alcohol and Tobacco on the Web: New Threats to Youth.* Washington, Center for Media Education.

Willis GB (2004). Cognitive interviewing revisited: A useful technique, in theory? In: Pressler S, Rothgeb JM, Couper MP, et al., eds., *Methods for Testing and Evaluating Survey Questionnaires.* Hoboken, NJ, Wiley & Sons: 23-44.

Willis, GB (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA, Sage Publications.

Willis GB, Lessler J (1999). *The BRFSS-QAS: A guide for systematically evaluating survey question wording*. Rockville, Maryland, Research Triangle Institute.

Wilson DH, Wakefield MA, Esterman A, et al. (1987). 15's: they fit in everywhere—especially the school bag: a survey of purchases of packets of 15 cigarettes by 14 and 15 year olds in South Australia. *Commun Health Stud*, 11(1 Suppl):16s-20s.

Wilson N, Thomson G, Howden-Chapman P, et al. (2006). Regulations should ban the sale of cigarette pack covers of health warnings. *N Z Med J*, 119(1243):U2251.

Witte K, Allen M (2000). A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Educ Behav*, 27(5):591-615.

Wong CM, Hu ZG, Lam TH, et al. (1999). Effects of ambient air pollution and environmental tobacco smoke on respiratory health of non-smoking women in Hong Kong. *Int J Epidemiol*, 28(5):859-864.

Woodruff S, Agro AD, Wildly MB (1995). Point of purchase tobacco advertising: Prevalence, correlates and brief intervention. *Health Values*, 19:56-62.

Wu W, Zhang L, Jain RB, et al. (2005). Determination of carcinogenic tobacco-specific nitrosamines in mainstream smoke from U.S.-brand and non-U.S.-brand cigarettes from 14 countries. *Nicotine Tob Res*, 7(3):443-451.

Wynder EL, Graham EA (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. *JAMA*, 143(4):329-336.

Yan M. (2007). Test-retest reliability and predictive validity for selected questions in the ITC Four Country Survey. ITC Project technical report. (*http://www. itcproject org)*.

Yoshimura K (2000). The psychological characteristics of tobacco dependence in a rural area of Japan. *J Epidemiol*, 10(4):271-279.

Young D, Borland R, Hammond D, et al. (2006). Prevalence and attributes of roll-your-own smokers in the International Tobacco Control (ITC) Four Country Survey. *Tob Control*, 15(Suppl 3):iii76-iii82.

Yurekli A.(2002). *Design and Administration: Design and Administer Tobacco Taxes*. Tool 4. World Bank Economics of Tobacco Toolkit. The International Bank for Reconstruction and Development. Washington, DC, the World Bank.

Zhang P (2002). *Tobacco Control: Understand and Evaluate the Impact of Tobacco Control Policies on Employment*. Tool 5. World Bank Economics of Tobacco Toolkit. The International Bank for Reconstruction and Development. Washington, DC, the World Bank.

Zuckerman M, Ball S, Black J (1990). Influences of sensation seeking, gender, risk appraisal, and situational motivation on smoking. *Addict Behav*, 15(3):209-220.

Appendix 1.
Fagerström Test for Nicotine Dependence (FTND) and Heaviness of Smoking (HSI)*

Please answer the following questions:

1. **How soon after you wake up do you smoke your first cigarette?**
   - 3 - Within 5 minutes
   - 2 - 6-30 minutes
   - 1 - 31-60 minutes
   - 0 - After 60 minutes

2. **Do you find it difficult to refrain from smoking in places where it is forbidden (e.g. in church, at the library, cinema, etc.)?**
   - 1 - Yes
   - 0 - No

3. **Which cigarette would you hate to give up?**
   - 1 - The first one in the morning
   - 0 - All the others

4. **How many cigarettes/day do you smoke?**
   - 0 - 10 or less
   - 1 - 11-20
   - 2 - 21-30
   - 3 - 31 or more

5. **Do you smoke more frequently during the first hours after waking than during the rest of the day?**
   - 1 - Yes
   - 0 - No

6. **Do you smoke if you are so ill you are in bed most of the day?**
   - 1 - Yes
   - 0 - No

* The Heaviness of Smoking Index (HSI) consists of FTND Item 1 and FTND Item 4, using the same response scales and calculating the total score using the sum of the scores on those two items.

Total score = Sum of all questions

Appendix 2.
Features of Diagnostic and Statistical Manual-IV (DSM-IV) Substance Dependence that are Targeted by Structured Diagnostic Interviews

A maladaptive pattern of substance use, leading to clinically significant impairment or distress as manifested by three (or more) of the following occurring at any time in the same 12-month period:

1. Tolerance, as defined by either of the following:
    a. A need for markedly increased amounts of the substance to achieve intoxication or desired effect.
    b. Markedly diminished effect with continued use of the same amount of substance.

2. Withdrawal, as manifested by either of the following:
    a. The characteristic withdrawal syndrome for the substance
    b. The same (or a closely related) substance is taken to relieve or avoid withdrawal symptoms.

3. The substance is often taken in larger amounts or over a longer period than was intended.

4. There is a persistent desire or unsuccessful efforts to cut down or control substance use.

5. A great deal of time is spent in activities necessary to obtain the substance (e.g. visiting multiple doctors or driving long distances), use the substance (e.g. chain smoking), or recover from its effects.

6. Important social, occupational, or recreational activities are given up or reduced because of substance use.

7. The substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance (e.g. current cocaine use despite recognition of cocaine-induced depression, or continued drinking despite recognition that an ulcer was made worse by alcohol consumption).

Appendix 3.
Features of the International Statistical Classification and Related Health Problems-10 (ICD-10) Substance Dependence that are Targeted by Structured Diagnostic Interviews

Three or more of the following manifestations should have occurred together for at least one month, or if persisting for periods of less than one month, should have occurred together repeatedly within a 12-month period:

1. A strong desire or sense of compulsion to take the substance.

2. Impaired capacity to control substance-taking behaviour in terms of onset, termination or level of use, as evidenced by: the substance being often taken in larger amounts or over longer periods of time than intended, or any unsuccessful effort or persistent desire to cut down or control substance use.

3. A physiological withdrawal state when substance use is reduced or ceased, as evidenced by the characteristic withdrawal syndrome for the substance, or use of the same (or closely related) substance with the intention of relieving or avoiding withdrawal symptoms.

4. Evidence of tolerance to the effects of the substance, such that there is a need for markedly increased amounts of the substance to achieve intoxication or desired effect, or that there is a markedly diminished effect with continued use of the same amount of the substance.

5. Preoccupation with substance use, as manifest by: important alternative pleasures or interests being given up or reduced because of substance use, or a great deal of time being spent in activities necessary to obtain the substance, take the substance, or recover from its effects.

6. Persisting with substance use despite clear evidence of harmful consequences as evidenced by continued use when the person was actually aware of the nature and extent of harm.

Appendix 4.
The Tobacco Dependence Screener (TDS)*

Please answer the following questions either yes or no:

1. Have you often had periods of days when you smoked a lot more than you than you intended to?

2. Have you ever tried to quit or cut down on tobacco and found you could not?

3. Did you crave tobacco after you quit or cut down on it?

4. Did you have any of the following problems when you quit or cut down on tobacco: irritation, nervousness, restless, trouble concentrating, headache, drowsiness, upset stomach, heart slow down, increased appetite or body weight, hand shakes, or depression?

5. Did you ever start using tobacco again to keep from having such problems?

6. Have you ever continued to smoke when you had a serious illness that you knew made it unwise to use tobacco?

7. Have you ever continued to use tobacco after you knew that it caused you health problems?

8. Did you continue to use tobacco after you knew that it caused you mental problems?

9. Have you ever felt like you were dependent on tobacco?

10. Have you given up work or social activities so you could use tobacco?

* To get the total score for the TDS, add up all the points by giving each “yes” response one point, and each “no” response zero points.

Appendix 5.
The Cigarette Dependence Scale (CDS)

1. Please rate your addiction to cigarettes on a scale of 0 to 100:†
   a. I am NOT addicted to cigarettes at all = 0
   b. I am extremely addicted to cigarettes =100
      1 - 0-20
      2 - 21-40
      3 – 41-60
      4 – 61-80
      5 – 81-100

2. On average, how many cigarettes do you smoke per day?†
      1 – 0-5
      2 – 6-10
      3 – 11-20
      4 – 21-29
      5 – 30+

3. Usually, how soon after waking up do you smoke your first cigarette?†
      5 – 0-5 minutes
      4 – 6-15 minutes
      3 – 16-30 minutes
      2 – 31-60 minutes
      1 – 61+ minutes

4. For you, quitting smoking for good would be:†
      5 – Impossible
      4 – Very difficult
      3 – Fairly difficult
      2 – Fairly easy
      1 – Very easy

Please indicate whether you agree with each of the following statements:
      1 – Totally disagree
      2 – Somewhat disagree
      3 – Neither agree nor disagree
      4 – Somewhat agree
      5 – Fully agree

5. After a few hours without smoking I feel an irresistible urge to smoke.†
6. The idea of not having any cigarettes causes me stress.
7. Before going out, I always make sure that I have cigarettes with me.
8. I am a prisoner of cigarettes.
9. I smoke too much.
10. Sometimes I drop everything to go out and buy cigarettes.
11. I smoke all the time.
12. I smoke despite the risks to my health.

The CDS total scores are sums of all of the relevant 5 or 12 items.

† Items included in the CDS-5.

Appendix 6.
The Nicotine Dependence Syndrome Scale (NDSS)

Circle the number that indicates how well each of the following statements describes you:

1 – Not at all true
2 – Somewhat true
3 – Moderately true
4 – Very true
5 – Extremely true

1. After not smoking for while, I need to smoke to relieve feelings of restlessness and irritability.
2. Whenever I go without a smoke for a few hours, I experience craving.
3. After not smoking for a while, I need to smoke in order to keep myself from experiencing any discomfort.
4. When I'm really craving a cigarette, it feels like I'm in the grip of some unknown force that I cannot control.
5. I feel a sense of control over my smoking. I can "take it or leave it" at any time.
6. I tend to avoid restaurants that don't allow smoking, even if I would otherwise enjoy the food.
7. Sometimes I decline offers to visit with my non-smoking friends because I know that I'll feel uncomfortable if I smoke.
8. Even if traveling a long distance, I'd rather not travel by airplane because I wouldn't be allowed to smoke.
9. Since the time when I became a regular smoker, the amount I smoke has either stayed the same or has decreased somewhat.
10. Compared to when I first started smoking, I need to smoke a lot more now in order to get what I want out of it.
11. Compared to when I first started smoking, I can smoke much, much more now before I start to feel nauseated or ill.
12. It's hard to estimate how many cigarettes I smoke per day because the number often changes.
13. My smoking pattern is very irregular throughout the day. It is not unusual for me to smoke many cigarettes in an hour, then not have another one until hours later.
14. The number of cigarettes I smoke per day is often influenced by other factors – how I'm feeling, what I'm doing, etc.
15. I smoke at different rates in different situations.
16. My smoking is not much affected by other things. I smoke about the same amount whether I'm relaxing or working, happy or sad, alone or with others, etc.
17. My cigarette smoking is fairly regular throughout the day.
18. I smoke consistently and regularly throughout the day.
19. I smoke about the same amount on weekends as on weekdays.

Scoring for the NDSS involves multiplying the item score by a factor loading score and then summing the factor-corrected scores for each subscale and for the total scale. See Shiffman *et al.* (2004) for the factor loadings.

Appendix 7.
Wisconsin Inventory of Smoking Dependence Motives (WISDM)

Below are a series of statements about cigarette smoking. Please rate your level of agreement for each, using the following scale:

1 = Not true of me at all
2
3
4
5
6
7 = Extremely true of me

1. I enjoy the taste of cigarettes most of the time.
2. Smoking keeps me from gaining weight.
3. Smoking makes a good mood better.
4. If I always smoke in a certain place it is hard to be there and not smoke.
5. I often smoke without thinking about it.
6. Cigarettes control me.
7. Smoking cigarettes improves my mood.
8. Smoking makes me feel content.
9. I usually want to smoke right after I wake up.
10. Very few things give me pleasure each day like cigarettes.
11. It's hard to ignore an urge to smoke.
12. The flavor of a cigarette is pleasing.
13. I smoke when I really need to concentrate.
14. I can only go a couple hours between cigarettes.
15. I frequently smoke to keep my mind focused.
16. I rely upon smoking to control my hunger and eating.
17. My life is full of reminders to smoke.
18. Smoking helps me feel better in seconds.
19. I smoke without deciding to.
20. Cigarettes keep me company, like a close friend.
21. Few things would be able to replace smoking in my life.
22. I'm around smokers much of the time.
23. There are particular sights and smells that trigger strong urges to smoke.
24. Smoking helps me stay focused.
25. Smoking helps me deal with stress.
26. I frequently light cigarettes without thinking about it.
27. Most of my daily cigarettes taste good.
28. Sometimes I feel like cigarettes rule my life.
29. I frequently crave cigarettes.
30. Most of the people I spend time with are smokers.
31. Weight control is a major reason why I smoke.
32. I usually feel much better after a cigarette.
33. Some of the cigarettes I smoke taste great.
34. I'm really hooked on cigarettes.
35. Smoking is the fastest way to reward myself.
36. Sometimes I feel like cigarettes are my best friends.
37. My urges to smoke keep getting stronger if I don't smoke.
38. I would continue smoking, even if it meant I could spend less time on my hobbies and other interests.
39. My concentration is improved after smoking a cigarette.
40. Seeing someone smoke makes me really want a cigarette.
41. I find myself reaching for cigarettes without thinking about it.
42. I crave cigarettes at certain times of the day.
43. I would feel alone without my cigarettes.

Appendix 7.
Wisconsin Inventory of Smoking Dependence Motives (WISDM)

44. A lot of my friends or family smoke.
45. Smoking brings me a lot of pressure.
46. Cigarettes are about the only thing that can give me a lift when I need it.
47. Other smokers would consider me a heavy smoker.
48. I feel a strong bond with my cigarettes.
49. It would take a pretty serious medical problem to make me quit smoking.
50. When I haven't been able to smoke for a few hours, the craving gets intolerable.
51. When I do certain things, I know I'm going to smoke.
52. Most of my friends and acquaintances smoke.
53. I love the feeling of inhaling the smoke into my mouth.
54. I smoke within the first 30 minutes of awakening in the morning.
55. Sometimes I'm not aware that I am smoking.
56. I'm worried that if I quit smoking I'll gain weight.
57. Smoking helps me think better.
58. Smoking really helps me feel better if I've been feeling down.
59. Some things are very hard to do without smoking.
60. Smoking makes me feel good.
61. Smoking keeps me from overeating.
62. My smoking is out of control.
63. I consider myself a heavy smoker.
64. Even when I feel good, smoking helps me feel better.
65. I reach for cigarettes when I feel irritable.
66. I enjoy the sensations of a long, slow exhalation of smoke.
67. Giving up cigarettes would be like losing a good friend.
68. Smoking is the easiest way to give myself a lift.

WISDM Subscale Scores = Mean of all subscale items

WISDM Total Score = Sum of all the subscale means

WISDM Subscale	Items
Affiliative Attachment	#20, 36, 43, 48, 67
Automaticity	#5, 19, 26, 41, 55
Loss of Control	#6, 28, 34, 62
Behavioral Choice/Melioration	#10, 21, 35, 38, 46, 49, 68
Cognitive Enhancement	#13, 15, 24, 39, 57
Craving	#11, 29, 37, 50
Cue exposure/Associative Process	#4, 17, 23, 40, 42, 51, 59
Negative Reinforcement	#7, 18, 25, 32, 58, 65
Positive Reinforcement	#3, 8, 45, 60, 64
Social/Environmental Goals	#22, 30, 44, 52
Taste/Sensory Process	#1, 12, 27, 33, 53, 66
Tolerance	#9, 14, 47, 54, 63
Weight Control	#2, 16, 31, 56, 61

Appendix 8.
The Fagerström Test for Nicotine Dependence-Smokeless Tobacco (FTND-ST)

1. How soon after you wake up to do you place your first dip?

Within 5 min	3
6–30 min	2
31–60 min	1
After 60 min	0

2. How often do you intentionally swallow tobacco juice?

Always	2
Sometimes	1
Never	0

3. Which chew would you hate to give up most?

The first one in the morning	1
Any other	0

4. How many cans/pouches per week do you use?

More than 3	2
2–3	1
1	0

5. Do you chew more frequently during the first hours after awakening than during the rest of the day?

Yes	1
No	0

6. Do you chew if you are so ill that you are in bed most of the day?

Yes	1
No	0

Appendix 9.
Quantitative Measures of Constructs to Assess Labelling Policies

I. Quantitative Measures

Health Warnings - Awareness

Measure	"Are you aware of any recent changes to health warnings on cigarette packs?" (Yes, No)
Source	Borland & Hill, 1997b
Outcome	Almost universal awareness among adult smokers in Australia.
Measure	"Have you noticed any changes to the health warnings on cigarette packages?" (Yes, No)
Source	Health Canada, 2001
Outcome	Almost universal awareness among general population in Canada, including non-smokers and youth.
Measure	"Have you ever seen health warning messages on cigarette packages?" (Yes, No)
Source	Hammond *et al.*, 2003
Outcome	Almost universal awareness among adult smokers in Canada.
Measure	"Have you noticed any changes to the warning labels on cigarette packs since [6 month anchor]?" (Yes, No) "Does the pack you are currently smoking have the new warnings?" (Yes, No)
Source	The ITC Project
Outcome	Used to evaluate the implementation of new UK warnings in 2003; high levels of awareness.
Measure	"Have you seen the new warning labels which include pictures?" (Yes, No, Don't know)
Source	Koval *et al.*, 2005
Outcome	Young adults: Current smokers and experimental/ex-smokers were more likely to have seen new pictorial warning labels than never-smokers.

Appendix 9.
Quantitative Measures

Health Warnings - Looking/Reading	
Measure	"About how often do you find yourself looking at, or reading health warning messages on cigarette packages?" (Never, Less than once a week, About once a week, Once every 2 or 3 days, About once a day, Several times per day)
Source	Health Canada, 2005
Outcome	Increased significantly following the implementation of new pictorial warnings.
Measure	"In the last month, that is, since [date], how often, if at all, have you noticed the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often) "In the last month, how often, if at all, have you read or looked closely at the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often)
Source	Hammond *et al.*, 2007a
Outcome	Measures of noticing and reading strongly associated with the size and comprehensiveness of warnings among Canadian, USA, UK, and Australian adult smokers. Changes in the warnings were associated with increases in noticing and reading in the UK.
Depth of Processing	
Measure	"In the past 3 months, how carefully have you ever read the inside messages in cigarette packs?" (5-point Likert scale) "In the past 3 months, how carefully have you ever read the outside messages in cigarette packs?" (5-point Likert scale) "In the past 3 months, how often have you thought about what the inside warnings have to say?" (5-point Likert scale) "In the past 3 months, how often have you thought about what the outside warnings have to say?" (5-point Likert scale) "In the past 3 months, have you ever talked about the new warning labels with other smokers or non-smokers?" (Never, Rarely, Sometimes, Often, All the time) "In the past 3 months, have you ever thought about the warning labels or what they had to say when a cigarette pack wasn't in sight?" (Never, Rarely, Sometimes, Often, All the time) "In the past 3 months, have you ever saved or held on to a warning label after you had finished the pack?" (Yes, No)
Source	Hammond *et al.*, 2004a
Outcome	Depth of Processing scale consisting of these measures was associated with intention to quit (cross-sectional analyses), as well as future cessation-related behaviour (decreases in consumption, attempt to quit, or abstinence) at 3-month follow-up, adjusting for demographics, intentions to quit, and measures of consumption.

Appendix 9.
Quantitative Measures

Health Warnings - Discussions with Others	
Measure	"Did the box encourage you to talk about smoking with other people?" (Never, Sometimes, Often) "Over the past 4 weeks, have you discussed smoking with other people?" (Never, Sometimes, Often)
Source	Christie & Etter, 2004
Outcome	After four weeks using cigarette pack covers with health warnings, almost one third (32%) said that the boxes often prompted discussions about smoking with others, 51% responded sometimes, and 16% said never.
Measure	"How often have people you know mentioned or discussed the new warnings on cigarette packs in conversations with you?" (Frequently, Sometimes, Rarely, Never)
Source	Canadian Cancer Society, 2001
Outcome	More than 80% of people had people they know discuss the new warnings.
Health Warnings - Media Sources	
Measure	"In the last 6 months, have you noticed advertising or information that talks about the dangers of smoking, or encourages quitting in any of the following places?" (Yes, No to a list of 9 sources, including "on cigarette packages")
Source	Hammond *et al.*, 2006a
Outcome	Between country differences observed: noticing information on cigarette packs was strongly associated with the size and strength of the warning in Canada, USA, UK, and Australia. Package warnings were the second most common source of health information after television.
Emissions - Looking/Reading	
Measure	"Overall, how often do you find yourself looking at, or reading, the information about chemicals and substances on the side of cigarette packages?" (Never, Less than once a week, About once a week, Once every 2 or 3 days, About once a day, Several times per day)
Source	Health Canada, 2003
Outcome	Descriptive: approximately 43% reported "never" looking at the information on the side of packages, whereas a quarter reported looking at the side once per week or more often.

Appendix 9.
Quantitative Measures

Measure	"In the last month, how often have you read or looked closely at the information about the contents on the side of the pack?" (Never, Rarely, Sometimes, Often, Very often)
Source	The ITC Project
Outcome	Descriptive: approximately 43% reported "never" looking at the information on the side of packages, whereas a quarter reported looking at the side once per week or more often. More than one half reported using the higher number in the range, mainly because it was "most harmful."
Health Warnings – Eye Tracking	
Measure	Eye tracking: Participants wore eye-tracking equipment and viewed USA cigarette advertisements with health warnings.
Source	Fischer *et al.*,1989b
Outcome	Average attention to warning was 8% of viewing time; the health warning was not viewed at all in almost half of all cases (44%). Viewing time associated with subsequent recall/recognition of health warnings.
Measure	Eye tracking: Participants wore eye-tracking equipment and viewed cigarette ads with health warnings, including existing mandated warnings in the USA and newly developed warnings.
Source	Krugman *et al.*, 1994
Outcome	The new warnings were more likely to attract attention, attract attention in a shorter period of time, although were less likely to hold attention over time.
Health Warnings – Viewing Time	
Measure	Health warnings were flashed on a screen and the amount of time was recorded.
Source	Peters *et al.*, 2007
Outcome	Longer viewing times were associated with picture warnings compared to text warnings.
Health Warning - Location	
Measure	"Where on the cigarette packages have you seen warning labels?" (Presented with diagram) "Circle all of the real warnings that you have actually seen on packages of cigarettes." (Four actual and four false)
Source	Robinson & Killen, 1997
Outcome	Increased knowledge of pack warnings associated with higher levels of smoking.

Appendix 9.
Quantitative Measures

Measure	"Without looking at a cigarette package, where on the pack are the warnings or messages located?" (Open ended)
Source	Hammond *et al.*, 2004a
Outcome	Participants showed good recall of outside warnings; lesser recall of inside warnings.
Measure	"Where are the warnings on Canadian cigarette packages located?" (Open ended)
Source	Environics Research Group, 2003
Outcome	Participants showed good recall of outside warnings; lesser, though still high, recall of inside warnings.
Measure	Knowledge of the presence and location of health warnings on packages.
Source	Richards *et al.*, 1989
Outcome	67% knew the warnings were on the side of the pack (91% of current smokers versus 60% of non-smokers).
Health Warning – Content	
Measure	"As far as you know, what do the health warnings on cigarette packets say?" (Open ended)
Source	Hill, 1988
Outcome	86% knew at least one health warning. 97% of smokers could provide text of a warning; smokers more knowledgeable about warning content. Knowledge of warnings may be associated with intention to quit.
Measure	Smokers were asked about the content of US Surgeon General's warnings on cigarette packages.
Source	Richards *et al.*, 1989
Outcome	Very few (7%) knew there were four different warnings. Content knowledge was low: 22% no knowledge, 48% knew general theme (health), 28% knew one specific theme, 1% knew wording for one. Smokers and non-smokers had similar results.
Measure	"Circle all of the real warnings that you have actually seen on packages of cigarettes." (Four actual and four false)
Source	Robinson & Killen, 1997
Outcome	Increased knowledge of pack warnings associated with higher levels of smoking.
Measure	"As far as you know, what do the health warnings on the front of cigarette packs say?" (Open ended)
Source	Borland & Hill, 1997a
Outcome	Increase in knowledge following implementation of more comprehensive policy.

Appendix 9.
Quantitative Measures

Measure	Students were asked to list everything they could remember about a cigarette package after they had viewed an image for approximately one minute.
Source	Rootman *et al.*, 1995
Outcome	Students in Canada were more likely to recall the health warning on Canadian packages (83%) than USA students were to recall warnings on USA packages (6%).
Measure	"I'm now going to describe some warning labels or messages that may or may not be on cigarette packages. I'd like you to tell whether you remember seeing each on packs, by answering yes or no." (Recognition: four actual, four false warnings) "Which of the following types of information are provided either on the outside or the inside of cigarette packages?" (Recognition: seven actual, one false) "Can you recall any specific quit-tips that appear on cigarettes packs?" (Open ended)
Source	Hammond *et al.*, 2004a
Outcome	Respondents provided a range of responses. The "mouth cancer" warning was the most common response.
Measure	"In your own words, write or describe the health warnings you remember." (Open ended)
Source	Health Canada Youth Smoking Survey (Brown *et al.*, 2005)
Outcome	Respondents provided a range of responses. "Mouth cancer" and "impotence" most commonly recalled warnings.
Measure	"Without looking at a cigarette package, what specific health warning messages can you remember seeing on cigarette packages in Canada?" (Open ended)
Source	Health Canada, 2003
Outcome	Respondents provided a range of responses. The "mouth cancer" warning was the most common response.
Measure	Participants were asked to identify current USA labels (Score out of four)
Source	O'Hegarty *et al.*, 2006
Outcome	Descriptive only: approximately half identified at least three of the four warning messages on USA cigarette packs.
Emission Side Panel - Content	
Measure	"Without looking at anything, what, if any, chemicals or substances can you name that are in cigarettes or cigarette smoke?" (Open ended)

Appendix 9.
Quantitative Measures

	"Without looking at a cigarette package, as far as you know, are any chemicals or substances currently listed on cigarette packages in Canada" (Yes, No)
	"Without looking at a cigarette package, can you name any chemicals or substances that are currently listed on cigarette packages in Canada?" (Open ended)
Source	Health Canada, 2003
Outcome	Higher recall for nicotine (64%) and tar (53%) than the four other emissions listed on packages (<25%). Daily smokers more likely to recall other emissions.
Measure	"Without looking at a pack, can you tell me the tar level of your cigarettes?" (Open ended)
Source	O'Connor *et al.*, 2006c
Outcome	Very few were able to correctly recall tar level. Smokers living in a country where the tar numbers were listed on packs were more likely to report the tar level.
Measure	"Can you tell me, in milligrams, the tar content of your cigarettes?" (Open ended)
	Smokers were asked where they could obtain information on the yield of the cigarette brand they smoked. (5 point scale: Very low (1-3mg), Low (4-6mg), Medium (7-9mg), High (10-12mg), Very high (10-12mg))
Source	Chapman *et al.*, 1986
Outcome	Only 2% of smokers correctly recalled the ISO tar level and a majority underestimated the level of their own brand.
Measure	"What is the tar number of the cigarettes you smoked most recently?" (Open ended) "Is a [5mg/16mg] tar cigarette lower in tar than most cigarettes on the market?" (Yes, No)
Source	Cohen, 1996b
Outcome	Few smokers knew the tar level of cigarettes, with the exception of those who smoked cigarettes in the 1-5mg FTC tar range.
Health Warnings – Affective Reactions	
Measure	"Some people have reported that the warning labels have made them feel different types of emotion. On a scale from 1 to 5 where, 1 is not at all and 5 is extremely, have the warning labels made you feel: fearful, amused, disgusted, angry?"
Source	Hammond *et al.*, 2004a
Outcome	Respondents who reported greater negative emotional responses were more likely to engage in cessation-related behaviour (i.e. attempts to quit, reductions in consumption, or abstinence) at 3-month follow-up.

Appendix 9.
Quantitative Measures

Measure	Response to smoking-related image or word cues on four adjective pairs (e.g. good-bad, positive-negative, favorable-unfavorable, and like-dislike) "How does this warning label make you think and feel about cigarette smoking?" on a 9-point scale (–4 = extremely negative to +4 = extremely positive)
Source	Peters *et al.*, 2007
Outcome	Canadian labels produced more negative affective reactions to smoking cues and to the smoker image, among both smokers and nonsmokers, without signs of defensive reactions from smokers. Participants in the Canadian label condition reported that their warning labels made them feel more negative toward smoking than those in the US label condition.
Health Warnings – Avoidance	
Measure	"Since the beginning of the year, have you ever concealed the warning messages on your cigarette package, either by placing a cardboard sleeve or other cover over your package, OR by transferring your cigarettes to another container?" (Yes, No for each option) "Do you currently do this with your cigarettes all the time, occasionally, rarely, or never?"
Source	Canadian Cancer Society, 2001
Outcome	Descriptive only.
Measure	"I try my best to avoid thinking about the warning labels." (Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree) "Have you made any efforts to avoid the labels by: (1) covering or hiding the labels? (2) using another case? (3) any other method?" (Yes, No to each question) "Have you ever bought another brand or requested a specific package to avoid a particular warning label?" (Yes, No)
Source	Hammond *et al.*, 2004a
Outcome	Approximately 40% reported at least one avoidance behaviour. Avoidance was not associated with future cessation related behaviour measured at 3-month follow-up.
Measure	"In the last month, have you made any effort to avoid looking at or thinking about the warning labels: (1) by covering the warnings up? (2) by keeping the pack out of sight? (3) by using a cigarette case or some other pack? (4) by not buying packs with particular labels?" (Yes, No to each question)
Source	The ITC Project
Outcome	Descriptive only.

Appendix 9.
Quantitative Measures

Health Warnings – Accuracy	
Measure	"How accurately do you feel the warnings depict the risks to your health?" (Very inaccurately, Somewhat inaccurately, Neutral, Somewhat accurately, Very accurately)
Source	Hammond *et al.*, 2004a
Outcome	Fewer than 15% of smokers reported that the information in the pictorial warnings was at all inaccurate.
Measure	"The messages are accurate." (Strongly Disagree, Somewhat disagree, Somewhat agree, Strongly Agree) "The messages provide you with important information about the health effects of smoking cigarettes." (Strongly Disagree, Somewhat disagree, Somewhat agree, Strongly Agree)
Source	Health Canada, 2005
Outcome	Descriptive only: Fewer than 10% of adults or youth disagreed that the warnings were accurate, while approximately 20% or less disagreed that the messages provide important information about health risks.
Health Warnings – Believability	
Measure	Credibility: 7 point bi-polar scale (informative-uninformative).
Source	Loken & Howard-Pitney, 1988
Outcome	Specific warnings on US cigarette advertisements were rated as credible.
Measure	"In your opinion, are each of the following sources of information about the chemicals and substances in cigarettes and cigarette smoke very, somewhat, not very, or not at all trustworthy ...? (1) Canadian Cancer Society, (2) Health Canada, (3) Tobacco companies
Source	Health Canada, 2003
Outcome	Well respected, non-governmental organisations and Health Canada were found to be highly credible sources of health information, whereas the tobacco companies were not.
Measure	"How much do you believe the information in the warning label is true or false?" on a 9 point scale (–4 = completely false to +4= completely true). US participants were asked whether Canadian labels should be used in the USA.
Source	Peters *et al.,* 2007
Outcome	No differences in the believability of text or graphic warnings. A majority of both smokers and nonsmokers endorsed the use of Canadian labels in the USA.

Appendix 9.
Quantitative Measures

Measure	"Do you believe the health warnings that you see on cigarette packages?" (Yes, No, Not sure, I haven't see them)
Sources	Health Canada Youth Smoking Survey, 2002; Brown *et al.,* 2005 (http:// www.hcsc.gc.ca/hl-vs/pubs/tobac-tabac/yss-etj-2002/index-eng.php)
Outcome	Almost universal agreement among youth that the health warnings were believable.
Measure	Perceived Believability Scale: Unbelievable/Believable, Untrustworthy/Trustworthy, Not convincing/Convincing, Not credible/Credible, Unreasonable/Reasonable, Dishonest/Honest, Questionable/Unquestionable, Inconclusive/Conclusive, Not authentic/Authentic, Unlikely/ Likely (Adjective pairs rated on 1-5 Likert scale)
Source	Beltramini, 1988
Outcome	Respondent's smoking behaviour (and demographics) had no effect on perceive believability of USA health warnings.
Measure	Beltramini's 10-item Perceived Believability Scale (see above).
Source	Cecil *et al.*, 1996
Outcome	Smokers score lower than non-smokers when viewing heath warnings.
Health Warnings – Public Opinion/Support	
Measure	Respondents were asked about the adequacy of current warnings, approval for more information if it meant that less youth would smoke, and approval of "rules to make cigarette packets less colourful and attractive." (Open ended)
Source	Borland & Hill, 1997a
Outcome	Descriptive only: Half thought adequate, a third thought there should be more - 88% approval if caused less youth to smoke - 60% for less attractive; 87% for less attractive, if reduced uptake
Measure	"How much do you agree or disagree with cigarette packages having health warning messages?" (Agree a lot, agree a little, Neither agree nor disagree, Disagree a little, Disagree a lot)
Source	Health Canada Youth Smoking Survey, 2002 (http://www.hc-sc.gc.ca/hl-vs/pubs/tobac-tabac/yss-etj-2002/index-eng.php)
Outcome	Descriptive only: high levels of support from youth smokers and non-smokers.

Appendix 9.
Quantitative Measures

Measure	"Would you like to see more or less of the following information on cigarette packages?" (More, Less, About right) 1. health risks 2. how to quit 3. benefits of quitting 4. where to get help to quit 5. 1-800 telephone # for info and advice 6. website address
Source	Hammond *et al.*, 2004a
Outcome	The majority of smokers reported a desire for more information for each variable. Fewer than 30% expressed a desire for less health information on packages.
Measure	Participants were asked their opinions about the size of the US labels. (Open ended)
Source	O'Hegarty *et al.*, 2006
Outcome	A higher percentage of former smokers than current smokers (62.0% and 40.8%, respectively) thought that current US labels should be larger.
Emission Labelling - Public Support	
Measure	"Cigarette manufacturers are currently required to list three chemicals - carbon monoxide, tar, and nicotine, and their amounts on cigarette packages. What do you think about requiring cigarette manufacturers to add to this list three other chemicals that are found in tobacco - formaldehyde, benzene, and hydrogen cyanide, and their amounts?" (Strongly support, Somewhat support, Somewhat oppose, Strongly oppose)
Source	Health Canada, 2001
Outcome	Descriptive only: approximately 90% of the general population indicated support, with approximately 80% of youth and adult smokers indicating support.
Measure	Participants were asked whether they agreed or disagreed that tar yields should be displayed wherever cigarettes are purchased. (Agree, Disagree, Unsure)
Source	Chapman *et al.*, 1986
Outcome	72% agreement.

Appendix 9.
Quantitative Measures

Health Warnings – Thinking About Health Risks	
Measure	"In the past 3 months, how have the warning labels affected how much you think about the health risks of smoking? Have they made you think about health risks: A lot less, A little less, No difference, A little more, A lot more?"
Source	Hammond *et al.*, 2004a
Outcome	Associated with intentions to quit cross-sectionally, as well as cessation-related behaviour at 3-month follow-up when combined with measures of depth of processing.
Measure	"To what extent, if at all, do the warning labels make you think about the health risks of smoking?" (Not at all, A little, Somewhat, A lot)
Source	Hammond *et al.*, 2007a
Outcome	Respondents living in countries with larger, more comprehensive warnings were more likely to report that the warnings made them think about the health risks of smoking. Changes in the UK warnings were also associated with increases in thinking about the health risks of smoking.
Measure	"Have the new health warnings made you think a lot more about the health effects of smoking, think a little more, or have they had no impact on how much you think about the health effects of smoking?"
Source	Canadian Cancer Society, 2001
Outcome	Descriptive only: approximately half of smokers and non-smokers reported thinking more about health risks because of the warnings.
Measure	"This [Canadian] label would make me more worried about the health effects of smoking." (5-point Likert scale where 5=strongly agree)
Source	O'Hegarty *et al.*, 2006
Outcome	Graphics were rated as more likely to cause worry about the health effects of smoking than text warnings.
Measure	"Do you agree or disagree that this warning is likely to prompt people to think more about the effects of [targeted health risk] on [target group]?" (1-Strongly disagree, 2-Disagree, 3-Neither agree nor disagree, 4-Agree, 5-Strongly agree, 6-Don't know)
Source	BRC Marketing & Social Research, 2004
Outcome	Question was used to evaluate message targeting similar themes (e.g. the risks of smoking while pregnant).

Appendix 9.
Quantitative Measures

Health Warnings - Concern & Worry About Health Effects	
Measure	"Have the new health warnings made you much more concerned about the health effects of smoking, a little more concerned, or have they had no impact?"
Source	Canadian Cancer Society, 2001
Outcome	Descriptive only: approximately 40% of smokers and non-smokers reported thinking more about health risks because of the warnings.

Health Warnings - Knowledge of Health Effects & Perceived Risk	
Measure	"Thinking about the health warning messages you have seen on cigarette packages, have these messages been very effective, somewhat effective, not very effective, or not at all effective in each of the following ways ... Informing you about the health effects of cigarette smoking? (Not at all effective, Not very, Somewhat, Very Effective)
Source	Health Canada, 2005
Outcome	A substantial proportion of smokers reported that the pictorial warnings were effective in informing them about the health effects of smoking.
Measure	"I am going to read you a list of health effects and diseases that may or may not be caused by smoking cigarettes. Based on what you know or believe, does smoking cause the following: (1) heart disease in smokers, (2) stroke in smokers, (3) impotence in male smokers, (4) lung cancer in smokers, (5) lung cancer in nonsmokers from secondhand smoke, (6) blindness, (7) mouth and throat cancer, (8) peripheral vascular disease, (9) asthma in children from secondhand smoke." (Yes, No to each question) Note: Not all health effects included in every wave.
Source	Hammond *et al.*, 2006a
Outcome	Specific health effects were associated with health effects listed on the label in each country.
Measure	"In your opinion, are there any illnesses caused by smoking?" If yes, "Which illnesses are caused by smoking? (Open ended) Smoking knowledge and attitudes (16 items)
Source	Borland & Hill, 1997b
Outcome	Smokers reported a greater number of smoking illnesses following implementation of new text warnings in Australia. Acceptance of statements used in warnings became stronger at follow-up.

Appendix 9.
Quantitative Measures

Measure	Risk scores for smoking, environmental tobacco smoke, susceptibility to lung cancer, respiratory diseases, and cardiovascular diseases, reduced life expectancy, and others.
Source	Portillo & Antonanzas, 2002
Outcome	Students attributed a higher health risk to smoking following the presentation of the EU warnings packages.
Measure	Cigarettes cause cancer. Cigarettes cause strokes and heart disease. Tobacco smoke causes fatal lung disease in nonsmokers. (5-pint Likert scale where 5= strongly agree)
Source	O'Hegarty *et al.*, 2006
Outcome	Significantly higher endorsement for two of the three statements following presentation of graphic versus text only warnings following presentation of the warnings.
Measure	"I am going to read you a list of human health effects and diseases that may or may not be caused by smoking cigarettes. Based on what you know or believe, please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree that smoking cigarettes can cause each of the following ... lung cancer, throat cancer, mouth cancer, emphysema, heart disease, asthma, premature death, chronic bronchitis, gum or mouth diseases, smaller babies/reduced growth of babies during pregnancy, stroke, wrinkles and premature ageing, premature birth or preterm birth, blood clots, miscarriages, stomach ulcers, impotence in men, infertility, bladder cancer, gangrene, acne, multiple sclerosis, hepatitis, arthritis, Alzheimer's disease." Note: a list of 11 health effects for secondhand smoke was also used.
Source	Health Canada, 2005
Outcome	Descriptive only
Emissions - Comprehension & Meaning	
Measure	"What in your opinion is the meaning of the tar value of cigarettes?" (Open ended) "Is a 10-mg tar cigarette more relevant to health than a 5-gm one, and if so, how much more?" (Yes, No; Open ended)
Source	Gori, 1990
Outcome	Approximately half reported that tar levels were an indicator of health risk. Overall, very low understanding of tar levels.

Appendix 9.
Quantitative Measures

Measure	"Could a pack-a-day smoker significantly lower health risks by switching from a 20-mg/16mg tar cigarette to a 5-mg tar cigarette?" (Yes, No)
	"Assume a person switched from a 10-mg tar cigarette to a 1-mg tar cigarette. Which of the following is closest to your opinion? The person probably could smoke more than 1, but these numbers can't tell you how much less tar the person would take in from the 1-mg tar cigarette. The person could smoke more than 1 or 2, but fewer than 9 or 10, of the 1-mg tar cigarette without taking in more tar. The person could smoke about 10 of the 1-mg tar cigarettes without taking in more tar."
Source	Cohen, 1996a,b
Outcome	Substantial minority of respondents reported that lower tar cigarettes would lower health risk or result in lower tar exposure.
Measure	"Tar numbers [appear/used to appear] in advertisements and sometimes on cigarette packs. As you understand it, how closely, if at all, are the tar numbers related to the amount of tar that smokers take into their bodies?" (Closely related, Somewhat related, Not at all related)
	"As far as you know, are each of the following chemicals included in cigarette smoke? (1) cyanide (2) mercury (3) arsenic (4) carbon monoxide." (Yes, No to each question)
	"Which of the following, if any, helps to indicate whether a cigarette brand COULD be less harmful compared to others: …. The tar or nicotine levels for a brand?" 1 A little less harmful 2 No different 3 A little more harmful
Source	The ITC Project
Outcome	Knowledge of chemicals was associated with labeling policy among smokers in Canada, the USA, UK, and Australia: if the emission was printed on the package, participants were more likely to report it was in smoke.
Measure	"Which of the following do you think is closest to the total number of chemicals or substances that are found in cigarettes or cigarette smoke? Is the total number closest to (3, 6, 15, 500, 1000, 4000, 5000)?"
	"Here are questions about some of the chemicals that are listed on the cigarette packs. What specific health effects, if any, can you name that can be caused by…(Each of 6 chemicals on side panel of package: tar, nicotine, CO, benzene, formaldehyde, hydrogen cyanide)?" (Open ended)
	"A range of numbers is reported beside each chemical on the side of the cigarette pack. For example, a pack may say "Tar 13 to 31mg." What does this range mean?" (Open ended)
	"Do you think the range of numbers listed for a chemical on the pack means ...?" (All cigarettes in that pack will have the same amount of a chemical, but those in another pack of the same brand may have more or less. Some cigarettes in that pack may have larger amounts of a chemical and others in the pack may have less. Some smokers may take in larger amounts of a chemical and other smokers may take in less. Combination of the above.)

Appendix 9.
Quantitative Measures

	"Now, still thinking about the numbers that go with the chemicals that are listed on the side of a cigarette package, have you frequently, sometimes, rarely, or never done each of the following ...?" Talked about/compared amounts with another smoker. Used amounts to inform about health hazards of own/other brand. Used amounts to look for brand that may be less harmful. Used amounts to look for/try another brand close to own. Used amounts as step to quit smoking. "If you were to look for a safer or less harmful cigarette, do you think you would or would not use the information about the amounts of chemicals listed on the cigarette packs to help you find a less harmful brand?" (Yes or Maybe, No, None less harmful)
Source	Health Canada, 2003
Outcome	Generally, low knowledge of health effects and very little understanding of what the range of numbers on Canadian cigarette packages means. Nevertheless, over half indicated they would use the emission information to identify a "less harmful" cigarette brand.
Light & Mild Descriptors - Health	
Measure	"Compared with smoking regular cigarettes, would smoking light cigarettes increase, decrease, or have no effect on your risk of having health problems?" "Is that GREATLY increase [decrease] or SOMEWHAT increase [decrease]?" "If the number 100 stood for the risk to health from a regular cigarette, and 1 stood for the risk to health for a nonsmoker, what number stands for the risk to the health of a smoker of light cigarettes?"
Source	Kozlowski *et al.*, 2000
Outcome	The numerical "1-100" approach was found to be misleading relative to the "ordered categorical" approach.
Measure	"How many light cigarettes would someone have to smoke to get the same amount of tar as from one regular cigarette?" (Open ended – respondent to provide number of cigarettes, or also could respond "don't know") "Now I'm going to ask you about reasons some people might give for smoking [light or ultra-light, according to self-reported usual type] cigarettes. For each one, please tell me whether it is one of your reasons for smoking [light or ultra-light] cigarettes. Do you smoke [light or ultra-light] cigarettes as a step toward quitting smoking completely? Do you smoke [light or ultra-light] cigarettes to reduce the risks of smoking without having to give up smoking? Do you smoke [light or ultra-light] cigarettes to reduce the tar you get from smoking? Do you smoke [light or ultra-light] cigarettes to reduce the nicotine you get from smoking? Do you smoke [light or ultra-light] cigarettes because you prefer the taste compared to regular cigarettes?" If the response were yes to any of these reasons, smokers were asked: "How important is this reason to you? Is it very important or somewhat important?"

Appendix 9.
Quantitative Measures

Source	Kozlowski *et al.*, 1998b
Outcome	The majority of smokers reported that lights would deliver lower amounts of tar and nicotine than regular cigarettes - a misconception.
Measure	Health knowledge summative score (from 8/10 items in 1996/2000 respectively) Perceptions of light/mild cigarettes Reasons for smoking light/mild
Source	Ashley *et al.*, 2001
Outcome	Approximately one quarter of smokers said they smoked lights to reduce health risks, 40% replied to smoke light/mild as a step toward quitting, and 41% said they would be more likely to quit if they knew that light cigarettes provided the same amount of tar and nicotine as regular cigarettes.
Measure	Respondents were asked whether light/ultra-light cigarettes in comparison to regular cigarettes were safer, healthier, and less likely to cause cancer. (5 point scale ranging from 1 = “definitely not true” to 5 = “definitely true”) Respondents were asked to estimate the number of light and ultra-light cigarettes, respectively, someone would have to smoke to get the same amount of tar in one regular cigarette. Respondents asked to estimate the risk of smoking lights and ultra-lights, respectively, relative to the risk of not smoking (designated “0”) and the risk of smoking regulars (designated “10”).
Source	Shiffman *et al.*, 2001
Outcome	On average, smokers believed that lights afforded a 25% reduction in risk, and ultra-lights a 33% reduction in risk. Light and ultra-light cigarette smokers evaluated the risks of their own cigarette types more favourably. On average, half of all smokers thought that it was necessary to smoke two light cigarettes and three ultra-light cigarettes to get as much tar as from a single regular cigarette. Believing that lights and ultra-lights delivered less tar and nicotine independently contributed to the belief that these cigarettes were safer.
Measure	“In your opinion, how many (a) light and (b) ultra-light cigarettes would someone have to smoke to inhale the same amount of nicotine as from one regular cigarette?” (Open ended)
Source	Etter *et al.,* 2003c
Outcome	On average, participants reported one would have to smoke two light cigarettes or four ultra-light cigarettes to inhale the same amount of nicotine from one regular cigarette.

Appendix 9.
Quantitative Measures

Measure	Smokers were exposed to print advertisements for light and regular cigarettes and asked to rank the products on health risk, amount of tar, and carcinogenicity, and identified the messages they perceived the advertisements to convey. (Rating scale from 1-10)
Source	Hamilton *et al.*, 2004
Outcome	Respondents perceived lights as having significantly lower health risks and carcinogen levels than regular cigarettes.
Measure	"The next question is about the amount of tar smokers take into their lungs from smoking cigarettes. Compared to smokers of regular cigarette brands, do smokers who smoke [participant's brand] take in: a lot less tar into their lungs than smokers of regular cigarettes, a little less, about the same amount, a little more tar, a lot more tar into their lungs?" "For the following questions, I will refer to all types of light, mild, and low tar cigarettes as "light cigarettes." Please tell me if you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements about light cigarettes: …Light cigarettes are less harmful than regular cigarettes. …Smokers of light cigarettes take in less tar than smokers of regular cigarettes." "How many light cigarettes would you have to smoke to harm you as much as 10 regular cigarettes would?" (Far fewer light cigarettes than 10, Somewhat fewer, Same number of light cigarettes, Somewhat more, Far more light cigarettes than 10) "Do you think that the brand you usually smoke, [current brand], might be a little less harmful, no different, or a little more harmful, compared to other cigarette brands? 1 A little less harmful 2 No different 3 A little more harmful "Which of the following, if any, helps to indicate whether a cigarette brand COULD be less harmful compared to others: ….Words in the name of the brand, such as light or mild?" 1 A little less harmful 2 No different 3 A little more harmful
Source	The ITC Project
Outcome	A majority of smokers surveyed in each country, except Canada, continue to believe that light cigarettes offer some health benefit compared to regular cigarettes (Canada 43%, USA 51%, Australia 55%, UK 70%). A majority of smokers in all four countries believed that light cigarettes are smoother on the throat and chest than regular cigarettes. Predictors of use of light cigarettes and beliefs about possible benefits were very similar in the four countries.

Appendix 9.
Quantitative Measures

Measure	Which of the following do you think is true: a light cigarette has more tar than a regular one, a light cigarette has less tar, or a light cigarette has the same amount of tar as a regular "Which of the following do you think is true: a light cigarette has more nicotine than a regular one, a light cigarette has less nicotine, or a light cigarette has the same amount of nicotine as a regular?" "If you switched to [light/regular] cigarettes, how do you think this would affect your daily intake of nicotine?" (Increase, Decrease, Remain same, Depends on cigarette) "If you switched to [light/regular] cigarettes, how do you think this would affect your daily intake of tar?" (Increase, Decrease, Remain same, Depends on cigarette)
Source	Castrucci & Gerlach, 2007
Outcome	The majority of smokers say that lights have less tar and/or nicotine, ultra-light smokers more likely to say these have less. In addition, 63.0% of light and 73.0% of ultra-light smokers reported that switching would increase their intake of tar and nicotine.
Light & Mild Descriptors - Sensory Properties	
Measure	"When you smoke a cigarette, is it easy or difficult to tell if it is a regular strength variety or a light one, just from the experience of smoking it?" (Open ended) "Light cigarettes are smoother on your throat and chest than regular cigarettes." (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree)
Sources	The ITC Project; Borland *et al*., 2004
Outcome	The majority of smokers contacted in Australia, Canada, UK, and USA believe light cigarettes are smoother on their throat and chest than regular cigarettes.
Measure	3-item Sensation index: "You cough less smoking lights." "Lights feel smoother on your throat." "Lights feel easier on your chest."
Source	Shiffman *et al*., 2001
Outcome	Believing that lights and ultra-lights were less harsh independently contributed to the belief that these cigarettes were safer.

Appendix 9.
Quantitative Measures

Light & Mild Descriptors - Addiction	
Measure	"For the following statement/question, I will refer to all types of light, mild, and low tar cigarettes as "light cigarettes." Please tell me if you strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with each of the following statements about light cigarettes: Light cigarettes make it easier to quit smoking. Do you believe that [light/ultra-light] cigarettes are more addictive, as addictive, or less addictive than regular cigarettes?"
Source	The ITC Project
Outcome	A minority of respondents reported that "light/mild" cigarettes may be less addictive.
"Other" Brand Descriptors	
Measure	"Which, if any, of the following terms on cigarette packs mean that the cigarettes are supposed to be some form of light, mild, or low-tar cigarette?" (Yes, No to each) 1 Smooth 2 Refined 3 Generous 4 Ultra "Do you think that the brand you usually smoke, [current brand], might be a little less harmful, no different, or a little more harmful, compared to other cigarette brands?" 1 A little less harmful 2 No different 3 A little more harmful
Source	The ITC Project
Outcome	None to date.
Attractiveness	
Measure	"How good is this advertisement?" (0-very bad, to 20-very good) "How familiar is this advertisement?" (0-very bad, to 20-very good) "Do you want to smoke a cigarette?" (-5-would hate to, to +5-very much indeed)
Source	Hyland & Birrell, 1979
Outcome	Presentation of a health warning increased desire to smoke. Presence of warning decreased perceived "goodness" of ad; did not affect perceived familiarity.

Appendix 9.
Quantitative Measures

Measure	Attractiveness scale: 7-point bipolar scale (attractive-unattractive).
Source	Loken & Howard-Pitney, 1988
Outcome	Specific warnings on cigarette advertisements can act as a counter-influence to an ad's appeal by making it appear less attractive and less persuasive than if the ad contained only a general warning, particularly for smokers.
Measure	Products shown to adolescents with/without warnings. "Would you ever use this product?" (6 point scale from "absolutely, definitely would not use it" to "absolutely, definitely would use it") "Would most kids your age use it?" (6 point scale from "absolutely, definitely would not use it" to "absolutely, definitely would use it")
Source	Brubaker & Mitby, 1990
Outcome	Less than half (43%) exposed to warnings recalled seeing them; a third of those who noticed the warnings recalled the message content.
Measure	"Do you think the new warnings make cigarettes packages look less attractive, more attractive, or has it made no difference to their attractiveness? "How often have you put your cigarette package away because you didn't want others to see the warning on the package? Have you done this?" (Often, Sometimes, Rarely, Never)
Source	Canadian Cancer Society, 2001
Health Warnings – Consumption Patterns	
Outcome	Over half of smokers (63%) reported that the warnings make cigarette packages look less attractive, and approximately one third of smokers reported that they prefer to purchase a pack without the new warnings.
Measure	"Are you less inclined or more inclined to purchase cigarettes that contain the new warnings?" "If, when buying cigarettes from a shop or a vending machine, you were able to choose between a pack with or without the new warnings, which one would you buy?"
Source	Willemsen, 2005
Outcome	Approximately one third of smokers reported that they prefer to purchase a pack without the new warnings; 14% became less inclined to purchase cigarettes because of the new warnings.
Measure	Auction method: smokers placed separate bids on two packs of cigarettes; one with a text-only warning and the other with a graphic image of a smoker with cancer.

Appendix 9.
Quantitative Measures

Source	Thrasher *et al.,* 2007
Outcome	The pack with a graphic image had a mean attributed value which was 17% lower ($3.21 pesos) than the normal pack with the text warning, and this difference was consistent and statistically significant across sociodemographic groups, extent of smoking, quit attempts, and levels of perceived smoking risks.
Measure	"How often, if at all, have you been tempted to have a cigarette but decide not to because of the new warnings on the packs?" (Once, A few times, Many times, Never) "What impact have the new warnings had on your smoking behaviour inside your home? Have they motivated you to smoke much less inside your home, somewhat less, or have they had no impact?"
Source	Canadian Cancer Society, 2001
Outcome	One fifth of smokers indicated that the warnings had stopped them from having a cigarette, and approximately one quarter reported smoking less in the home as a result of the warnings.
Measure	"Thinking about the health warning messages you have seen on cigarette packages, have these messages been very effective, somewhat effective, not very effective, or not at all effective in each of the following ways ... Getting you to smoke less around others over the past year than you used to. Getting you to smoke less this year than last year."
Source	Health Canada, 2005
Outcome	Descriptive only: responses to all measures increased following implementation of larger pictorial warnings.
Measure	"In the past 3 months, have the warning labels made you smoke: a lot less, a little less, no difference, a little more, a lot more?" "In the past 3 months, have the warning labels ever made you delay before lighting up or butt out a cigarette early? (5 point Likert scale)
Source	Hammond *et al,* 2004a
Outcome	Approximately one fifth of Canadian smokers reported that the pictorial warnings had made them smoke less; less than 1% reported smoking more as a result of the warnings.
Measure	"In the last month, have the warning labels stopped you from having a cigarette when you were about to smoke one?" (Never, Once, A few times, Many times)

Appendix 9.
Quantitative Measures

Source	Hammond *et al.*, 2007a
Outcome	Larger pictorial warnings were associated with a greater likelihood of reporting forgoing a cigarette among Canada, USA, UK, and Australian smokers.
Measure	"Are you smoking (somewhat) less or (somewhat) more as a result of the new warnings or are you still smoking the same amount?"
Source	Willemsen, 2005
Outcome	Approximately 10% of adult smokers reported they smoked less because of the warnings.
Measure	Cigarettes smoked per week using data from national survey.
Source	Gospodinov & Irvine, 2004 (using data from the Canadian Tobacco Use Monitoring Survey)
Outcome	A reduction of 2 cigarettes per week among current smokers in the months following the implementation of pictorial health warnings.
Health Warnings - Smoking Initiation	
Measure	"Do you think the new warning labels might make some young people less likely to start smoking?" (Yes, No, Don't know) "Do you think the new warnings might make some young people more likely to start smoking?" (Yes, No, Don't know)
Source	Koval *et al.*, 2005
Outcome	Among young adults, current smokers were less likely than experimental/ex-smokers to believe that warning labels with stronger messages would make people their age less likely to smoke. Experimental/ex-smokers were more likely to believe that new warning labels would make people their age less likely to smoke than never- or current-smokers. Although only ~8% of current smokers were more likely to believe that new warning labels might make people their age more likely to smoke.
Health Warnings – Motivation to Quit	
Measure	"To what extent have the new warnings increased your motivation to quit smoking? Has your motivation increased: a lot, a little, not at all?"
Source	Canadian Cancer Society, 2001
Outcome	Descriptive only: approximately 40% reported the warnings had increased their motivation to quit.

Appendix 9.
Quantitative Measures

Measure	"Thinking about the health warning messages you have seen on cigarette packages, have these messages been very effective, somewhat effective, not very effective or not at all effective in… increasing your desire to quit smoking over the past year?"
Source	Health Canada, 2005
Outcome	None to date.
Measure	"How have the warnings affected the likelihood that you will quit smoking within the next year?" (A lot less likely to quit because of the labels, Somewhat less likely because of the labels, No difference, Somewhat more likely to quit because of the labels, A lot more likely to quit) "How have the warning labels affected your self-confidence in your ability to quit?" (A lot less confident in ability to quit, Somewhat less confident, No influence, Somewhat more confident, A lot more confident)
Source	Hammond *et al*, 2004b
Outcome	Approximately one third of smokers reported they were at least somewhat more likely to quit as a result of the pictorial warnings in Canada, and approximately one quarter reported that the warnings had made them more confident in their ability to quit.
Measure	"To what extent, if at all, do the warning labels on cigarette packs make you more likely to quit smoking?" (Not at all, A little, Somewhat, A lot) "In the past 6 months, have each of the following things led you to think about quitting? …warning labels" (Not at all, Somewhat, Very Much). Note: asked a part of a list.
Source	Hammond *et al*., 2007a
Outcome	Larger pictorial warnings were associated with greater proportions of smokers reporting that the warnings increased their likelihood of quitting among Canada, USA, UK, and Australian smokers.
Measure	"Did the new health warnings make you more or less motivated to quit smoking?"
Source	Willemsen, 2005
Outcome	Approximately 18% of Dutch smokers reported that new EU text warnings motivated them to quit.
Measure	"This label would motivate me to quit smoking." (5-point Likert scale, with 5=strongly agree)
Source	O'Hegarty *et al*., 2006

Appendix 9.
Quantitative Measures

Outcome	Respondents were significantly more likely to report that graphic warnings would motivate them to quit smoking compared to text warnings following presentation of the warnings.
Measure	"Do the new warnings make you think about trying to quit?" (Yes, No, Don't know)
	"In the past month, has noticing the new warnings led you to decide not to have a cigarette?" (Yes, No, Don't know)
Source	Koval *et al.*, 2005
Outcome	Young adults: ~40% of current smokers said new warnings made them think about trying to quit; ~25% said noticing warnings led them to not have a cigarette.
Health Warnings – Quit Attempts & Abstinence	
Measure	"Thinking about the health warning messages you have seen on cigarette packages, have these messages been very effective, somewhat effective, not very effective or not at all effective in… getting you to try to quit smoking within the past year?"
Source	Health Canada, 2005
Outcome	Descriptive only.
Measure	"To what extent, if at all, were the following reasons for your current quit attempt… warning labels?" (Not at all, Somewhat, Very much). Note: asked as part of a list of different reasons for quitting.
Source	Hammond *et al.*, 2007a
Outcome	Larger pictorial warnings were associated with greater proportions of smokers reported the warnings as a reason for their quit attempt among Canada, USA, UK, and Australian smokers.
Measure	Prevalence estimates for weekly smokers from national survey.
Source	Gospodinov & Irvine, 2004 (using data from the Canadian Tobacco Use Monitoring Survey)
Outcome	No discernable change in prevalence rates in the months following the introduction of pictorial warnings.
Health Warnings - Use of Cessation Services	
Measure	"What was the main reason for calling the quitline?" (Open ended)
Source	UK Department of Health
Outcome	UK pack warnings were the second largest reason cited by callers to the NHS Stop Smoking Helpline. Between 1,500 and 4,000 callers per month have cited this reason since the written warnings were introduced in 2003; a 12% increase.

Appendix 9.
Quantitative Measures

Measure	Call volume before and after introduction of quitline number on Dutch cigarette packages.
Source	Willemsen, 2002
Outcome	A 3- to 4-fold increase in call volume between the months before and after the new warnings.
Health Warnings - Quitting Among Former Smokers	
Measure	"How much did the warning labels on cigarette packages influence your decision to quit?" 1. No influence on your decision to quit 2. Very little influence on your decision to quit 3. Moderate influence on your decision to quit 4. Strong influence on your decision to quit 5. Main or major influence on your decision to quit "Did the warning labels make it easier or help you to quit?" 1. Not at all helpful 2. Only a little bit helpful 3. Moderately helpful 4. Very helpful 5. Extremely helpful
Source	Hammond *et al.*, 2003
Outcome	Asked along with price, bans/bylaws, personal health effects, health effects of others.
Measure	"To what extent, if at all, do the warning labels on cigarette packs make you more likely to stay quit?" (Not at all, a little, Somewhat, A lot)
Source	The ITC Project
Outcome	More prominent warnings associated with higher responses.
Measure	"To what extent have the new warnings on cigarette packages made you feel better about being a non-smoker? Have they made you feel a lot better, a little better, or have they had no impact on you?"
Source	Canadian Cancer Society, 2001
Outcome	Approximately half of former smokers reported that the warnings had made them feel better about being an ex-smoker.
Measure	"This label would motivate me not to start smoking again." (5-point Likert scale)
Source	O'Hegarty *et al.,* 2006
Outcome	Respondents were significantly more likely to report that graphic warnings would motivate them to remain abstinent compared to text warnings following presentation of the warnings.

Appendix 10.
Qualitative Measures from Focus Groups

Focus Groups - Health Warning Noticing & Salience	
Measure	"Does this warning catch your attention?" (Open ended) "Does it make you want to read further/know more?" (Open ended) "What stands out most to you?" (Open ended)
Source	Health Canada, 2006
Outcome	The picture was generally the first feature people looked at and related to; it determined the strength of the warning's emotional impact and noticeability. Pictures showing children, or clearly depicting disease (or diseased people) in some way, were the most effective. Motivation to read further varied based on the emotional impact of the warning itself and/or the personal relevance of the particular topic.
Measure	"Which graphics are most noticeable? Least noticeable? Why?" (Open ended) "Which are the most memorable and least memorable graphics? Why?" (Open ended) "Why are the warnings memorable?" (Open ended)
Source	Elliott & Shanahan Research, 2002
Outcome	Examined the content of images (e.g. shocking versus non-shocking, attractive versus unattractive). A variety of images and image styles is most likely to be effective in terms of maintaining "freshness" and retaining smoker attention.
Focus Groups - Health Warning Location	
Measure	"Can you describe what is displayed (shown) on a pack of cigarettes?" (Open ended) "What would you find when you look at a pack of cigarettes (without actually looking at a pack)?" "Can you describe all that is written on a cigarette pack?" "What do you recall about these warnings? What strikes you, what catches your attention?" "Now, think only of the images you remember having seen. Describe all the images you can recall." "Now, forget the images and think of only the words and what was written. Name all the words you can recall." "For each image recalled, ask: can you recall the words associated with this image?"
Source	CREATEC, 2003
Outcome	Descriptive only

Appendix 10.
Qualitative Measures from Focus Groups

Focus Groups - Health Warning Affective Reactions	
Measure	"Did you notice who made these warnings?" (Open ended) "Why do you think Health Canada made these warnings?" "Who else should make these warnings?"
Source	CREATEC, 2003
Outcome	Most thought the warnings came from the government.
Measure	"What do you think/how do you feel about this warning?" (Open ended) "What do you think/how do you feel about the picture?" "What do you think/how do you feel about the words?" "What does this warning tell you about the effects of smoking?" "As a smoker, does this warning affect you personally?"
Source	Health Canada, 2006
Outcome	The emotional impact of a warning appeared to predict its ability to inform and/or motivate thoughts of quitting. The most effective warnings generated a strong emotion supported by factual information.
Measure	Examined emotional reactions to warnings, including positive/negative message approach (e.g. positive could relate to feeling better by not smoking).
Source	Elliott & Shanahan Research, 2002
Outcome	Graphics had considerable impact on all age groups. Descriptive or emotive messages had considerable impact for younger smokers. Too much fear is likely to lead to defensiveness and rationalising of the messages; some warnings and explanatory messages need to provide support and encouragement.
Focus Groups - Health Warning Believability/Credibility	
Measure	"Are [the messages] truthful, personally relevant?" (Open ended; explore more with respondent)
Source	Elliott & Shanahan Research, 2002
Outcome	The relevance of the warnings depended upon the demographic of the smoker.
Measure	"Do you agree or disagree that any or all of these messages would be more effective being associated with or sponsored by the Ministry of Health?" (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree, Don't know) "For what particular reasons do you say that?" (Open ended)

Appendix 10.
Qualitative Measures from Focus Groups

Source	BRC Marketing & Social Research, 2004
Outcome	A large proportion of participants agreed messages would be more effective if they were associated with the Ministry of Health, as it gave official credibility.
Measure	“Do you believe what this warning is saying?” (Open ended)
Source	Health Canada, 2006
Outcome	While new information tended to interest participants, many also wanted proof or evidence in the form of statistics or clearer pictures. Lack of supporting data was often a key argument for rejection of disturbing new information. Most participants felt the Health Canada name lent credibility to the claim in the warning. Some participants tended to refute the message based on the idea that it was "not only" smoking that caused the illness or situation to occur.
Focus Groups - General Comprehension/Meaning	
Measure	Overall comprehension – are they easy to understand, is the information reliable? Any comprehension difficulties?
Source	Elliott & Shanahan Research, 2002
Outcome	Any increase in the font size, area of pack devoted to the message, and any contrasting background will facilitate readability. All photos and visuals need to be clear and recognizable to enable smokers to easily identify with the health issue concerned. Accompanying text messages need to be brief and as simple as possible to enable ease of comprehension.
Measure	“What message is this warning trying to get across?” (Open ended) ”Anything else it’s trying to say?” (Open ended) “What changes would you make to this warning to make it easier to understand?” (Open ended)
Source	Health Canada, 2006
Outcome	Pictures played the key role in understanding the message, and tended to override the meaning conveyed by the words in the headline. Some participants tended to take the words in the headline literally, and often failed to read in-between the lines or to derive an implicit message.

Appendix 10.
Qualitative Measures from Focus Groups

Measure	"Are [the warnings] interesting and informative? Helpful? Why/why not?" (Open ended) "How likely are [you] to read the explanatory messages? Is it curiosity? Information seeking?" (Open ended) "Do the labels raise the salience of health concerns?" (Open ended) "Which health topics/issues to do with smoking are smokers most concerned about? Why?" (Open ended)
Source	Elliott & Shanahan Research, 2002
Outcome	Health messages' impact increases with participant's age. Messages about children and babies effective in middle age range. Recommend including both factual and personalised
Measure	"Did you learn something while looking at these warnings? What?" (Open ended) "Are these warnings a good way to make you think? Why? Do they inform you?" (Open ended) "Do you take into account what is being said in the warnings?" (Open ended)
Source	CREATEC, 2003
Outcome	Descriptive only
Measure	"What does this warning tell you about the effects of smoking?" (Open ended) "Anything new here?" (Open ended) "After looking at these warnings, what do you remember about what you saw or read?" (Open ended) "Is there anything else?" (Open ended)
Source	Health Canada, 2006
Outcome	Overall, people's attitude towards new information was positive and was sometimes related to a warning's noticeability. If presented effectively (impactful picture and clear headline), most wanted more information.
Measure	Three standard readability tests: Flesch, Gunning's Fog, Dale/Chall
Source	Malouff *et al.*, 1992
Outcome	All three methods produced similar results: each of the four US warnings required a reading level typical of college students/graduates; the three smokeless tobacco warnings required middle/high school reading levels.
Measure	Participants were asked to look at their cigarette packages and instructed to offer what knowledge they had about each listed ingredient and how it can affect one's health. (Open ended)
Source	Health Canada, 2003
Outcome	Low knowledge of health effects

Appendix 10.
Qualitative Measures from Focus Groups

Focus Groups - Likelihood of Quitting	
Measure	"Do you agree or disagree that this packet (including the warning, picture and text) is likely to encourage [target group] to quit smoking or think about quitting?" (1-Strongly disagree, 2-Disagree, 3-Neither agree nor disagree, 4-Agree, 5-Strongly agree, 6-Don't know)
Source	BRC Marketing & Social Research, 2004
Outcome	Question was used to evaluate message targeting similar themes (e.g. the risks of smoking while pregnant).

Working Procedures for the IARC Handbooks of Tobacco Control

Starting in 2006, the series of International Agency for Research on Cancer (IARC) *Handbooks* of Cancer Prevention added tobacco control as a new area of prevention for their reviews. When appropriate, in addition to cancer, other health outcomes preventable by avoiding tobacco use may be included for evaluation in a *Handbook*.

The text that follows is organised in two principal parts. The first addresses the general scope, objectives and structure of the *Handbooks* of Tobacco Control. The second describes the scientific procedures for evaluating cancer-preventing agents or interventions.

The Working Procedures described herein are largely taken from the *Handbooks* of Cancer Prevention devoted to Chemoprevention and Screening, and from the IARC Monograph Preamble (updated in January 2006).

The term "exposure" appears repeatedly in these procedures, borrowed from the IARC *Monographs* devoted to the evaluation of carcinogenicity. Epidemiological studies conducted to assess the association between exposure to a given hazard and disease outcome are based on the meaning of the term "exposure" implying increased risk to an undesired health effect. However, in this series of *Handbooks* dedicated to the evaluation of the preventive effects of compounds, biological or pharmaceutical products, behaviours, programmes and interventions, the traditional meaning of the term "exposure" is unfitting. Therefore in several instances the term "intervention", which lacks a hazardous connotation, is preferred. Examples of interventions with expected benefits in the area of tobacco control are smoking cessation, banning of smoking in public places and taxation on cigarettes. The evaluation of their health effects may be the focus of future *Handbooks*.

Part one: General Principles

General Scope

The prevention and control of cancer are the strategic objectives of the International Agency for Research on Cancer. Cancer prevention may be achieved at the individual level by avoiding cancer-causing agents and at the population level by adopting programmes, legislation and regulations to reduce exposure to cancer-causing agents.

The *Handbooks* of Tobacco Control will evaluate the available evidence on the role of chemical compounds, biological and pharmaceutical products, behaviours, programmes and interventions in reducing tobacco use and decreasing tobacco-associated morbidity and mortality. The aim of the *Handbook* series is to provide the scientific community, policymakers and governing bodies of IARC member states as well as of other countries with evidence-based assessments of these interventions at the individual and population levels, with the ultimate goal of assisting in the global implementation of tobacco control provisions within national and international programmes aimed at reducing tobacco-related morbidity and mortality.

Objectives

The objective of the programme is to prepare, and to publish in the form of *Handbooks*, critical reviews and consensus evaluations of evidence on the preventive effect or risk reduction resulting from interventions focusing on tobacco control, with the help of an internationally formed Working Group of experts. The *Handbooks* may also indicate where additional research efforts are needed, specifically when data immediately relevant to an

evaluation are not available. The evaluations in the *Handbooks* are scientific and qualitative judgements of the peer-reviewed published data, conducted during a week-long meeting of peer review and discussions by the Working Group.

Topic for the Handbook

The topic to be evaluated in a *Handbook* is selected approximately twelve months prior to the meeting by the head of the Lifestyle, Environment and Cancer Group after consultation with IARC scientists involved in tobacco research. A *Handbook* may cover a single topic or a group of related topics in the area of Tobacco Control.

Meeting Participants

Soon after the topic of a *Handbook* is chosen, international scientists with relevant expertise are identified by IARC staff, in consultation with other experts. IARC uses literature searches to identify most experts. Each participant serves as an independent scientist and not as a representative of any organisation, government or industry.

Five categories of participants can be present at *Handbook* meetings: Working Group Members, Invited Specialists, Representatives of national and international health agencies, Observers and the IARC Secretariat. Participants in the first two groups generally have published significant research related to the topic being reviewed or in tobacco control in particular. Consideration is also given to demographic diversity and balance of area of expertise. All participants are listed, with their addresses and principal affiliations, at the beginning of each *Handbook* volume.

1. The *Working Group* is responsible for the critical reviews and evaluations that are developed during the meeting. The tasks of the Working Group are: (i) to ascertain that all appropriate data have been collected; (ii) to select the data relevant for the evaluation on the basis of scientific merit; (iii) to prepare accurate summaries of the data to enable the reader to follow the reasoning of the Working Group; (iv) to critically evaluate the results of epidemiological, clinical, and other type of studies; (v) to prepare recommendations for research and for public health action; and (vi) if the topic being reviewed so permits, to make an overall evaluation of the evidence of a protective effect or reduced risk associated with the exposure or intervention focus of the evaluation. Working Group members are selected based on knowledge and experience pertinent to the topic evaluated and absence of real or apparent conflicts of interest.

2. *Invited Specialists* are experts who also have critical knowledge and experience but have a real or apparent conflict of interest. These experts are invited when necessary to assist in the Working Group by contributing their unique knowledge and experience during subgroup and plenary discussions. They may also contribute text on the intervention being evaluated. Invited Specialists do not serve as meeting chair or subgroup chair, or participate in the evaluations.

3. *Representatives* of national and international health agencies may attend meetings because their agencies are interested in the topic of a *Handbook*. Representatives do not serve as meeting chair or subgroup chair, draft any part of a *Handbook*, or participate in the evaluations.

4. *Observers* with relevant scientific credentials may be admitted to a meeting by IARC in limited numbers. Priority will be given to achieving a balance of Observers from constituencies with differing perspectives. They are invited to observe the meeting and should not attempt to influence it. Observers serve as sources of first-hand information from the meeting to their sponsoring organisations. Observers also can play a valuable role in ensuring that all published

information and scientific perspectives are considered. Observers will not serve as chair or subgroup chair, draft any part of a *Handbook*, or participate in the evaluations. At the meeting, the chair and subgroup chairs may grant Observers the opportunity to speak, generally after they have observed a discussion.

5. The *IARC Secretariat* consists of scientists who have relevant expertise and who are designated by the Agency to attend a meeting. They serve as rapporteurs and participate in all discussions. When requested by the meeting chair or subgroup chair, they may also draft text or prepare tables and analyses.

The WHO Declaration of Interest form is sent to each prospective participant at the first contact, with the preliminary letter presenting the *Handbook* meeting. Before an official invitation is extended, each potential participant, including the IARC Secretariat, completes the WHO Declaration of Interests to report financial interests, employment and consulting, and individual and institutional research support related to the topic of the meeting. IARC assesses the declared interests to determine whether there is a conflict that warrants some limitation on participation. Working Group Members are selected based on the absence of real or apparent conflicts of interest. If a real or apparent conflict of interest is identified, then the expert is asked to attend as an Invited Specialist. The declarations are updated and reviewed again at the opening of the meeting, approximately 8 months later. Interests related to the subject of the meeting are disclosed to the meeting participants and in the published volume (Cogliano *et al*., 2004).

Data for the Handbooks

The *Handbooks* review all pertinent studies on the intervention to be evaluated. Only those data considered relevant to evaluate the evidence are included and summarized. Those judged inadequate or irrelevant to the evaluation may be cited but not summarized. If a group of similar studies is not reviewed, the reasons are indicated.

With regard to reports of basic scientific research, epidemiological studies and clinical trials, only studies that have been published or accepted for publication in the openly available scientific literature are reviewed. In certain instances, government agency reports that have undergone peer review and are widely available can be considered. Exceptions may be made ad hoc to include unpublished reports that are in their final form and publicly available, if their inclusion is considered pertinent to making an evaluation. Abstracts from scientific meetings and other reports that do not provide sufficient detail upon which to base an assessment of their quality are generally not considered.

Inclusion of a study does not imply acceptance of the adequacy of the study design or of the analysis and interpretation of the results, and limitations identified by the Working Group are clearly outlined in square brackets (ie, [ ]). The reasons for not giving further consideration to an individual study are also indicated in square brackets. Important aspects of a study, directly impinging on its interpretation, are brought to the attention of the reader. In general, numerical findings are indicated as they appear in the original report; units are converted when necessary for easier comparison. The Working Group may conduct additional analyses of the published data and use them in their assessment of the evidence. These analyses and their results are outlined in square brackets or in italics in the *Handbook*.

Working Procedures

(a) Literature to be reviewed

After the topic of the *Handbook* is chosen, pertinent studies are identified by IARC from recognized sources of information such as PubMed and made available to Working Group members and Invited Specialists to prepare the working papers for the meeting. Meeting participants are invited to supplement the IARC literature searches with their own searches. Studies cited in the working papers are available at the time of the meeting.

(b) Chair of the Meeting

The chair of the *Handbook* meeting is identified among leading international experts soon after the topic of a *Handbook* is chosen. The chair will help develop an outline for the *Handbook* early on, participate on conference calls with Working Group members and Invited Specialists in preparing for the meeting, provide early feedback on working papers and chair the meeting.

(c) Working papers

Working papers are due about 6 to 8 months after original contact of invited experts. The first version of the working papers is compiled and formatted by IARC staff about two months prior to the meeting, or as soon as they are received, and made available ahead of time through IARC's Internet to all Working Group members, Invited Specialists and the IARC Secretariat. Reception of working papers ahead of the established deadline is encouraged, as it allows review of their content, facilitating identification of information gaps early enough. When possible or when deemed necessary, some working papers may be discussed early on among experts to expedite the review process to be accomplished during the meeting. A conference call will be scheduled after reception of all working papers and prior to the meeting, with the aim of identifying areas deserving additional work by experts before the meeting.

Acknowledgement of significant contributions to the chapters by colleagues of the invited experts, either at their home institution or elsewhere, can be included in the *Handbook* under an acknowledgement paragraph to be shown following the listing of the meeting participants.

(d) Meeting

The Working Group members meet at IARC for seven to eight days to discuss and finalize the texts of the *Handbook* and to formulate the evaluations. The Working Group members and Invited Specialists are grouped into sub-groups according to their area of expertise. Sub-groups meet during the first three to four days to review in detail the first versions of their working papers, develop a joint subgroup draft, and write summaries. Scheduling of plenary and sub-group time may change from one *Handbook* meeting to another. During the last few days the participants meet in plenary session to review the subgroup working papers, summaries and to develop the consensus evaluations.

(e) Post-Meeting

After the meeting, the draft of the *Handbook* composed during the meeting is verified (by consulting the original literature), edited and prepared for publication by IARC staff. The aim is to publish *Handbooks* within twelve months of the meeting. If applicable, summaries reporting the results of the evaluation may be available on the IARC website (http://www.iarc.fr) soon after the meeting, and a short report may be published in the international literature.

Part two: Scientific Review of the Evidence and Evaluation

1. Scientific Review

The results of the studies reviewed will constitute the evidence forming the foundation of the evaluation. The validity of these studies should be examined critically to determine the weight of the studies contributing to the assessment. This will entail judging the appropriateness of study design, data collection (including adequate description of the intervention and follow-up), data analysis, and ultimately deciding if chance, bias, confounding or lack of statistical power may account for the observed results. The experts will ascertain how the limitations of the studies affect the results and conclusions reported. The criteria that follow apply to epidemiological and clinical studies and therefore may not be as relevant to studies where other quality criteria would be indicated—for example, those assessing the impact of economic policies.

(a) Quality of studies considered

It is necessary to take into account the possible roles of bias, confounding and chance in the

interpretation of epidemiological studies. Bias is the operation of factors in the study design or execution that lead erroneously to a stronger or weaker association than in fact exists between the exposure/intervention being evaluated and the outcome. Confounding is a form of bias that occurs when the association with the disease is made to appear stronger or weaker than it truly is as a result of an association between the apparent causal factor and another factor that is associated with either an increase or decrease in the incidence of the disease. The role of chance is related to biological variability and the influence of sample size on the precision of estimates of effect.

In evaluating the extent to which these factors have been taken into account in an individual study, the *Handbook* considers a number of aspects of design and analysis as described in the report of the study.

First, the study population, disease (or diseases) and exposure/intervention should have been well defined by the authors. Cases of disease in the study population should have been identified independently of the intervention of interest, and the intervention should have been assessed in a way that was not related to disease status.

Second, the authors should have taken into account—in the study design and analysis—other variables that can influence the risk of disease or impact of an intervention, and that may have been related to the intervention of interest. Potential confounding by such variables should have been dealt with either in the design of the study, such as by matching, or in the analysis, by statistical adjustment. In cohort studies, comparisons with local rates of the disease may or may not be more appropriate than those with national rates. Internal comparisons of disease frequency among individuals at different levels of the intervention are also desirable in cohort studies, since they minimize the potential for confounding related to difference in risk factors between an external reference group and the study population.

Third, the authors should have reported the basic data on which the conclusions are founded, even if sophisticated statistical analyses were employed. They should have given the numbers of exposed and unexposed cases and controls in a case-control study and the numbers of cases observed and expected in a cohort study. Further tabulations by time since exposure began and other temporal factors are also important. In a cohort study, data on all cancer sites and all causes of death should have been given to reveal the possibility of reporting bias. In a case-control study, the effects of investigated factors other than the exposure of interest should have been reported.

Finally, the statistical methods used to obtain estimates of relative risk, absolute rates of cancer, confidence intervals and significance tests, and to adjust for confounding should have been clearly stated by the authors. These methods have been reviewed for case-control studies (Breslow & Day, 1980) and for cohort studies (Breslow & Day, 1987).

Aspects that are particularly important in evaluating experimental studies are: the selection of participants, the nature and adequacy of the randomisation procedure, evidence that randomisation achieved an adequate balance between groups, the exclusion criteria used before and after randomisation, compliance with the intervention in the intervention group, and 'contamination' with the intervention in the control group. Other considerations are the means by which the end-point was determined and validated, the length and completeness of follow-up of the groups, and the adequacy of the analysis. Detailed analyses of both relative and absolute risks in relation to temporal variables, such as age at first exposure, time since first exposure, duration of exposure, cumulative exposure, peak exposure (when appropriate) and time since exposure ceased, will be reviewed and summarized when available.

Independent population-based studies of the same exposure or intervention may lead to ambiguous results. Combined analyses of data from multiple studies may be a means of resolving this ambiguity. There are two types of combined analysis: The first involves combining summary statistics such as relative risks from individual studies (meta-analysis), and the second involves a pooled

analysis of the raw data from the individual studies (pooled analysis).

The advantages of combined analyses include increased precision due to increased sample size as well as the opportunity to explore potential confounders, interactions and modifying effects that may explain heterogeneity among studies in more detail. A disadvantage of combined analyses is the possible lack of compatibility of data from various studies due to differences in subject recruitment, data collection procedures, measurement methods and effects of unmeasured covariates that may differ between studies.

Meta-analyses may be conducted by the Working Group during the course of preparing a *Handbook* and are identified as original calculations by placement of the results in square brackets or in italics. These may be de-novo analyses or updates of previously conducted analyses that incorporate the results from new studies. Whenever possible, however, such analyses are preferably conducted prior to the *Handbook* meeting. Publication of the results of such meta-analyses prior to or concurrently with the *Handbook* meeting is encouraged for purposes of peer review. The same criteria for data quality that would be applied to individual studies must be applied to combined analyses, and such analyses must take into account heterogeneity between studies.

(b) Criteria for causality

After the quality of each study has been summarized and assessed, a judgement is made concerning the strength of evidence that the exposure or intervention in question reduces the risk of disease or is protective for humans. Hill (1965) lists areas for evaluating the strength of epidemiological associations used in the review of human data when assessing carcinogenesis. These criteria, in many instances, will apply to the assessment included in a *Handbook*

- Consistency of observed associations across studies and populations;
- Magnitude of the reported association;
- Temporal relationship between exposure/intervention and change in disease;
- Exposure-response biologic gradient;
- Biological plausibility;
- Coherence of results across other lines of evidence; and
- Analogy present in related exposures and their effects on health.

If the results are inconsistent among investigations, possible reasons (such as differences in level of exposure/intervention) are sought, and results of studies judged to be of high quality are given more weight than those of studies judged to be methodologically less sound.

When several studies show little or no indication of an association between an intervention and cancer prevention, the judgement may be made that, in the aggregate, they show evidence of lack of effect. The possibility that bias, confounding or misclassification of exposure or outcome that could explain the observed results should be considered and excluded with reasonable certainty.

2. Summary of the data reviewed (evidence)

This section summarizes the results of the evidence presented in the preceding sections in a *Handbook* in a concise manner. Traditionally, this section does not include citation of literature as do preceding sections presenting and discussing the evidence covered in a *Handbook*.

3. Evaluation of the evidence

An evaluation of the strength of the evidence for disease prevention or reduction in morbidity and mortality is made using standard terms. It is conceivable that not every exposure/intervention reviewed in a *Handbook* of tobacco control will permit a formal evaluation of the evidence, as traditionally done in other *Handbooks* of Cancer Prevention and in the Monographs. In evaluating the strength of the evidence, a topic may allow a more formal evaluation (i.e. assigning causality or a protective effect in the prevention of cancer).

If assignment of causality is pertinent and possible, the possible outcomes of an evaluation can include:

Sufficient evidence of a reduction in risk: The Working Group considers that a causal relationship has been established between the intervention under consideration and a reduction in morbidity and mortality. That is, a relationship has been observed between the exposure/intervention and disease morbidity and mortality in studies in which chance, bias and confounding could be ruled out with reasonable confidence. A statement that there is *sufficient evidence* should be followed by a separate sentence that identifies the types of cancer and other diseases where a decreased morbidity and mortality was observed in humans.

Limited evidence of a reduction in risk:
An association has been observed between the exposure/intervention under consideration and a reduction in disease morbidity and mortality for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence.

Inadequate evidence of a reduction in risk: The available studies are of insufficient quality, consistency or statistical power to permit a conclusion regarding the presence or absence of a causal association between the exposure/intervention and a reduced morbidity and mortality. Alternatively, this category is used when no data are available.

Evidence suggesting lack of effect: There are several adequate studies that are mutually consistent in not showing an association between the exposure/intervention and disease morbidity and mortality. A conclusion of evidence suggesting *lack of risk reduction* is inevitably limited to the disease sites, conditions and levels of control, and length of observation covered by the available studies.

4. Overall evaluation

The overall evaluation, usually in the form of a narrative, will include a summary of the body of evidence considered as a whole and summary statements made about the strength of the evidence for a health protective or preventive effect, or adverse effects, as appropriate.

5. Recommendations

After reviewing the data and deliberating on them, the Working Group may formulate recommendations, where applicable, for further research and public health action.