

Cancer Incidence in Five Continents

Volume XI

Call for Data Data Specification Submission Notes

June 2015

Introduction

Since the first publication by Sir Richard Doll and colleagues in 1966, *Cancer Incidence in Five Continents* (CI5) has become an invaluable resource for cancer researchers and those involved in planning, monitoring and evaluating cancer control programmes worldwide. It provides a vital link between population-based cancer registries around the world, the [International Association of Cancer Registries](#) (IACR) and the [International Agency for Research on Cancer](#) (IARC).

The new Volume XI (CI5-XI) will include incident cancers for the period 2008-2012 and this document provides detailed instructions on the content and process of the submission. Consistent with previous practice, submitted datasets will be processed and analyzed by relevant IARC staff and stored on secure servers in order to ensure confidentiality. These datasets will then be carefully evaluated by the CI5-XI Editorial Board to ensure the incidence data published in CI5 are sufficiently comparable, complete and accurate (see also [Chapter 6 of Volume X](#)). As well as an abridged printed volume, a greater focus will be placed on developing new data exploration tools online from those presently available on the [CI5 web site](#) [CI5-X application].

For relevant registries, the datasets will also be used to update the annual time trend series available online, presently in CI5plus available [here](#). Please remember that, as an IACR publication, datasets will only be considered from IACR member registries and those who are not yet members should consult the IACR website (<http://www.iacr.com.fr/>) for application details.

The deadline for data submission will be 1 November 2015 and it is essential that this is respected.

Summary

This page summarizes the most important aspects of the CI5-XI call for data. Please click on the different links to access more detailed information about each element.

General Information

- The target incidence period for CI5-XI is **2008-2012**. Data from three consecutive years within this time period are the minimum that will be considered for publication.
- Cancer registries are, however, encouraged to submit all available data from years prior to 2008, and if complete, after 2012. The published version of volume XI will only contain 2008-2012 data but the pre-2008 information will be used to update previously supplied datasets and assess certain quality issues (see [Chapter 5](#) and [Chapter 6](#) of Volume X for additional information about the editorial and data processes). The post-2012 data will be included with other available years (if considered complete) in the next iteration of CI5plus, ensuring the series is as up-to-date as possible.
- The **deadline** for data submission is **1 November 2015**.

Checklist of Items Required

- **Incidence data** as a listing of individual cancer cases
- **Population data** from official censuses or intercensal/post-censal estimates
- **Mortality data** (where available), preferably from official vital statistics offices
- **A coding file** if the coding rules differ from those requested in this document
- **Completion of the online questionnaire** to provide details of the registry and the catchment population
- **An introductory text** describing the registry (part of the online questionnaire)

Data Quality and Coding

Incidence data should be verified and corrected prior to submission using, for example, [IARCcrgTools](#). Note that for some combinations of site/morphology/behaviour, accurate recoding requires a decision to provide corrections. Users can use the [IARC flag](#) variable to mark validated records.

File Format and Data Submission

Please submit separate files for incidence, population and mortality data. The datasets should be field-separated using any of the following: comma, semi-column, tab or pipe character (|). A fully automated and secured mechanism has been set up for submissions to CI5-XI at <https://cinportal.iarc.fr>. Where possible, all registries are requested to use this facility for the submission of files and for completion of the questionnaire. Files should not be zipped together but may be zipped individually (if they are large). They may be protected by a password which is then to be provided on the system.

Permission to Use Data

When submitting the data file through the above-mentioned registry portal, it is necessary to select the relevant boxes for the current call for data (i.e. **CI5-XI**; where relevant, **CI5plus**). **The boxes selected indicate that permission is granted to use the material submitted for those purposes.** As stated in the letter, all registry datasets, irrespective of inclusion in CI5-XI will be included in the relevant **continental report**. Please note that data in CI5-XI may also be used indirectly to develop estimates for **GLOBOCAN 2015**. Registries can be assured however that the submitted data will not be used for purposes other than those outlined above, without the explicit permission of the individual registry.

Please contact the editorial office (ci5@iarc.fr) for any questions.

Incidence Dataset

General Information

All primary tumours diagnosed at all ages between 2008 and 2012 should be submitted, including (if collected) basal cell and squamous cell carcinoma of the skin and non-malignant tumours of the central nervous system and of the urinary bladder. Data should be submitted as a case listing (one case per line).

Required Variables (mandatory variables in bold and preceded by a *****)([Read More](#)).

Variable	Format	Unknown/Missing	Definition/Notes
Ethnic group		99	If available, Read more
*Patient ID		Not allowed	Unique ID of the patient in the registry Read more
Tumour sequence #	00 = single tumour 01 = 1 st of several tumours 02 = 2 nd of several tumours...	99	If only one tumour is identified for a patient, this variable should contain "00". If several tumours are identified, numbering should start at "01". Read more
*Date of Birth	YYYY-MM-DD	9999-99-99	Read more
*Sex	1 = Male 2 = Female	9	
*Date of Incidence	YYYY-MM-DD	Not allowed	Read more
*Age in Years		999	Last completed year of age: <1 = 0, >99 = 100. Read more
*ICDO-3 Topography	ICDO-3 Definition (with letter C)	Not allowed	E.g. C531
*ICDO-3 Morphology	ICDO-3 Definition	Not allowed	E.g. 8170
*ICDO-3 Behaviour	ICDO-3 Definition	Not allowed	E.g. 3
*Basis of Diagnosis	ICDO-3 Definition	Not allowed	Read more
	0 = DCO 1 = Clinical 2 = Microscopic	9	Minimum coding requirements
Vital Status	1 = Alive 2 = Dead	9	Read more
Date of Death	YYYY-MM-DD	9999-99-99	
Date of Last Contact	YYYY-MM-DD	9999-99-99	Read more
IARC Flag	1 = OK 2 = Checked 0 = Failed	9	Read more

Definition of Variables

Ethnic group

If the registry data permits analysis by ethnic group, the relevant category should be appropriately coded in the Cancer Cases File AND the Population File using the same sub-populations. If mortality data are provided, they should also (where possible) include this variable. When information on ethnic group is provided, a description of the codes used must also be submitted.

Patient identification number

This is the unique number (or character string) permitting the registry identification of the record in their files. The patient identification number should be unique to ensure that if the same individual is reported again to the registry, s/he will be recognized as being the same person. This could also be a unique (national) personal identification number used for other purposes (birth register, insurance, etc.). In conjunction with the variable 'Tumour sequence number' it allows the editors to recognize multiple tumours as registered by the registry. **The submitted information should NOT allow the identification of a person in real life. Names of patients should never be included in the Cancer Cases File.**

Tumour sequence number

This number, allocated by the registry, permits distinction as well as order of occurrence of different tumours registered for the same patient. If only one tumour is identified for a patient, this variable should contain "00". If more than one tumour is identified, numbering should start at "01".

Dates

The calendar year in **all dates** should always be coded to four digits (with century: for example, 1903, 2003, etc.). All dates should be submitted in the following format: YYYY-MM-DD

Date of birth

If date of birth cannot be provided in full, the next possible/available detail should be included, such as YYYY-MM-99 if only the year and month of birth are available.

Date of incidence

Month and year are required as a minimum. Registries commonly define incidence date from one of a number of available dates and use hierarchical rules for determination of the date of incidence. Some of these rules are available for [download](#).

Age (in years)

The last completed year of age is to be used in the calculation. Cases less than one year of age should be coded as age 0. Do **not** use the same code for unknown age and over 99 years. Code for unknown age should be 999. If a different code is used, it must be described in the [Coding File](#).

Basis of diagnosis

ICD-O-3 coding is recommended. As a minimum, the information provided should include the following four crude categories so as to allow distinguishing of the method of cancer diagnosis:

- Death certificate only (DCO) (see definition in Volume X, Chapter 5)
- Clinical only
- Microscopically verified
- Unknown

Vital status

Together with “date of death” and “date of last contact”, information on vital status for each patient would facilitate survival analyses. Use of data in such analyses will not be made without consent / collaboration of each Registry, following distribution of a detailed protocol. These analyses are not part of the current CI5-XI proposal, but submission of vital status may be useful for setting up future projects.

Date of last contact

Date of last contact is the date when the patient was last confirmed to be either alive or dead. This is not the date when an attempt for contact was made, if the current vital status could not be confirmed.

IARC flag

Please use this variable to indicate which records have already been checked, for example using [IARCcrgtools](#). This would greatly reduce the processing time and avoid unnecessary request for verifications to your registry.

Population Dataset

General Information

The possible sources of population data are from official censuses, or from intercensal/postcensal estimates provided by Vital Statistics Departments or equivalent. Census data (with reference date) should be supplied for the period during, before and after the years covered in your dataset of cancer records. Any official estimates of the population made during the period should also be sent, preferably for each individual calendar year. If possible, population figures should give the mid-year (as of 1st July) estimates (or mid-period estimates) for each sub-category.

References to all sources of population data are requested and should be provided in the questionnaire. This information will be printed in the publication (or online) under the population pyramid of the registry page. Any unpublished estimates should be documented in the questionnaire. If population data are not available for this period, any available data should be submitted, clearly describing their source, reference area and period.

The population dataset, and the codes used in this file, should correspond to the cancer cases file with respect to ethnic groups, registration area, time period, sex and age range.

Population Data Variables

(mandatory variables in bold and preceded by a *).

Variable	Format	Unknown/Missing	Definition/Notes
*Year	4 digits, YYYY	Not allowed	
*Sex	Same as Incidence	Not allowed	
*Age / Age group		Not allowed	Age groups: 1 = 0-4 years of age, 2 = 5-9 years of age, 3 = 10-14 years of age, ..., 18 = 85+
*Number of residents		Not allowed	
Ethnic group	Same as Incidence	99	If provided in case listing

File Format

Each line of the Population File should include the number of residents for a combination of calendar year, sex and age. Age should be provided as a single year if possible or as standard 18 age-groups otherwise, using the same age groups codes as for the incidence data (e.g. 1=0-4 years of age, 2=5-9 years of age, ... 17=80-84 years of age, 18=85+). The coding of the age groups should be appropriately adjusted and documented if there are less than 18 age groups available. Please provide the numbers of persons of unknown age if applicable.

Examples of acceptable file formats:

Year	Sex	Age	Number of residents
1991	1	0	
1991	1	1	
1991	1	2	
1991	1	...	
1991	1	100	
2008	2	100	

Year	Sex	5-year age group	Number of residents
1991	1	1	
1991	1	2	
1991	1	3	
1991	1	...	
1991	1	18	
2008	2	18	

Mortality Dataset

General Information

The mortality data should consist of all certified deaths from cancer among residents of the registration area during the same period as that covered by the incidence data. **Wherever possible, the mortality data should be the official cancer mortality data**, as obtained from the Vital Statistics Department or equivalent and based on certificates/death records. Mortality data (where available) will be used in Volume XI, as in previous volumes, to calculate the ratios of mortality to incidence by site.

National cancer registries are **not** required to submit **national** mortality statistics, as these can be retrieved directly from the WHO database. The mortality data for **sub-national regions** covered by a cancer registry should include all past residents whose underlying cause of death was cancer.

The mortality dataset as well as the codes used in this file should correspond to the cancer cases file with respect to ethnic groups (if available), registration area, time period, sex, age range and cause of death.

Mortality Data Variables

(mandatory variables in bold and preceded by a *).

Variable	Format	Unknown/Missing	Definition/Notes
*Year	4 digits	Not allowed	
*Sex	Same as Incidence	Not allowed	
*Age / Age group		Not allowed	Age groups: 1 = 0-4 years of age, 2 = 5-9 years of age, 3 = 10-14 years of age, ..., 18 = 85+
*Cause of death	ICD-10	Not allowed	E.g. C61
*Number of deaths		Not allowed	
Ethnic group	Same as Incidence	99	If provided in case listing

File Format

Each line of the Mortality File should include number of deaths for a combination of calendar year, sex, age and cause of death (3 characters of the applicable ICD code, preferably ICD-10).

Age should be provided as a single year if possible or as standard 18 age-groups otherwise, using the same age groups codes as for the incidence data (e.g. 1=0-4 years of age, 2=5-9 years of age, ... 17=80-84 years of age, 18=85+). The coding of the age groups should be appropriately adjusted and documented if there are less than 18 age groups available. The total number of deaths can be provided if no breakdown information by age-group is available.

Examples of acceptable file formats:

Year	Sex	Age	Cause of Death	Number of deaths
1991	1	0	C00	
1991	1	1	C00	
1991	1	2	C01	
1991	1	
1991	1	100	C97	
2008	2	100	C97	

Year	Sex	5-year age group	Cause of Death	Number of deaths
1991	1	1	C00	
1991	1	2	C00	
1991	1	3	C01	
1991	1	
1991	1	18	C97	
2008	2	18	C97	

Coding File

This file is to be used to supply details of any non-standard coding conventions used or to provide information on any unique codes specific to your registry. Examples of the possible contents:

- Changes in the registration coverage
- Local definition of incidence date
- Coding of basis of diagnosis that differs from the [codes proposed in this document](#)
- Coding of ethnic group
- Description of the methodology used to derive population figures in the absence of published population data
- Any other information useful for processing the submitted dataset
- Any other information useful to evaluation of the results

Questionnaire

The questionnaire is an essential part of the submission. The information provided in the questionnaire will help the Editorial Board in their evaluation of the submitted datasets. Selected information from the questionnaire will be tabulated in the printed volume (or online) and will help those using published data to interpret results correctly. The questionnaire is available on the registries portal and should be completed online. Unless otherwise specified, all answers should refer to the target registration period, 2008 – 2012.

To access the questionnaire, please connect to the registries portal on <https://cinportal.iarc.fr> using the username and the password of your registry and select the 'questionnaires' section.

Introductory Text

The questionnaire includes a section for provision of a brief description of the registry. This should follow the outline provided in the questionnaire, and will be published in the printed volume (or online). The outline provides possible topics to be included. It is advisable to consult the introductory texts published in the previous volume of CI5. The objective is to help the reader interpret the incidence rates during the reference period and possibly any significant changes since previous publication in CI5.

If you cannot access the online questionnaire, please contact the editorial office (CI5@iarc.fr).

Downloads / Links

Publications

- **Cancer Incidence in 5 Continents, Volume X:** [Download](#)

Software

- **IARCcrgtools:** [Download](#)

Websites

- **International Association of Cancer Registries (IACR):** [Access](#)
- **International Agency for Research on Cancer (IARC):** [Access](#)
- **Cancer Incidence in Five Continents (Web Application):** [Access](#)
- **Registries' portal:** [Access](#)

Classifications and Coding

- **International Classification of Diseases for Oncology (ICD-O-3):** [Access](#)
- **European Network of Cancer Registries (ENCR) Recommendations for Coding of Incidence**
Date: [Download](#)
- **SEER Program Coding and Staging Manual:** [Download](#)
- **Multiple Primary rules:** [IARC/IACR](#), [SEER](#)
- **Rules for coding of "Basis of Diagnosis" (IARC):** [Download](#)

Other

- **Questionnaire in printable format:** [Download](#)