



# SURVCAN-3: Cancer Survival in Countries in Transition

Call for Data

Data specification

September 2016

**NB The deadline for data submission is 1 February 2017**

## SUMMARY

### General information

- SURVCAN-3 is the third initiative to produce reliable and comparable survival statistics in countries in transition, expanding the coverage area of previous projects.
- In addition to the provision of survival estimates for benchmarking purposes in countries in transition, SURVCAN-3 aims to enhance registries' expertise in collecting the local data needed to participate in this project, as well as support the local capacity to perform survival analyses in the registry.
- It will include all cancer diagnoses (ICD-10 codes C00-96) between 1 January 2006 and 31 December 2012, with a minimal follow-up of 2 years. Not all registries will be able to collect data for all cancer sites or for the entire diagnosis period; please contact us if this is the case.
- The deadline for data submission is **1 February 2017**.

This document provides detailed instructions on the content and process of submission. IARC will support the process of data collection and all data will be centrally processed at IARC using standard protocols and expert guidance, which will aid in improving the quality of the final data and results. The decision on accepting the individual registry data for survival analysis will depend upon the data quality indices.

### 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/behavior 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, mortality and life table 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>. 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 box for the current call for data (i.e. SURVCAN-3). The box selected indicates that permission is granted to use the material submitted for those purposes. Registries can be assured that the submitted data will not be used for purposes outside of the SURVCAN-3 framework without the explicit permission of the individual registry.

If you have any questions or concerns, please contact the editorial office: [Survival@iarc.fr](mailto:Survival@iarc.fr)

## 1. Incidence dataset

### 1.1. General Information

All primary incidence cancer cases (ICD-10 codes C00-96) registered during 2006-2012 with a potential follow-up of at least 2 years or followed until 31 December 2014 should be submitted as a case listing (one case per line), including (if collected) basal cell and squamous cell carcinoma of the skin and non-malignant tumors of the central nervous system and of the urinary bladder.

The details for each variable requested are detailed in the following pages.

**Table 1: Summary of data variables requested from registries for the incidence dataset**

Person-Related Data		Disease-Related Data		Follow-up Related Data	
Mandatory	Optional	Mandatory	Optional	Mandatory	Optional
Person Code	Socioeconomic group	Tumor sequence		Vital status	
Date of Birth	Ethnic group	IARC flag		Date corresponding to vital status	
Sex	Urban/Rural	Age at diagnosis		Type of follow-up	
		Incidence date		Method of follow-up	
		If ICD-O codes are used: - Topography - Morphology - Behavior			
		If ICD-10 codes are used: - Topography			
		Basis of diagnosis			
		Clinical extent of disease			
		TNM staging			

**Table 2: Description and format of all data variables for incidence dataset (mandatory variables preceded by a \*)**

Variable Name	Variable	Format	Unknown/Missing	Description
<b>*var1</b>	<b>Person Code</b>	Numeric or alphanumeric	Not allowed	Registry assigned unique number for each patient.
<b>*var2</b>	<b>Date of Birth</b>	YYYY-MM-DD	9999-99-99	At least year must be coded.  <i>If it is routine practice in your registry to substitute an imputed value for a missing value, it must be described in the coding file.</i>
<b>*var3</b>	<b>Sex</b>	1=Male 2=Female	9	
var4	Socioeconomic group	Numeric	99	See definition of variables.
var5	Ethnic group	Numeric	99	See definition of variables.
var6	Urban or rural setting	1=Urban 2=Rural	9	
<b>*var7</b>	<b>Tumor Sequence #</b>	0=Single tumor 1=1 <sup>st</sup> of multiple tumors 2=2 <sup>nd</sup> of multiple tumors	9	If a patient has only one tumor then this variable should be coded "0". If patient has multiple tumors then the numbering should start at "1".
<b>*var8</b>	<b>IARC flag</b>	0-4	9	See definition of variables.
<b>*var9</b>	<b>Age at diagnosis in years</b>	0-99	999	Last completed year.  <i>For different codes used for ages, it must be described in the coding file.</i>

Variable Name	Variable	Format	Unknown/Missing	Description
*var10	Incidence date	YYYY-MM-DD	9999-99-99	The hierarchal rules followed by the registry to determine the date of incidence as submitted for CI5 may be followed. At least year must be coded.  <i>If it is routine practice in your registry to substitute an imputed value for a missing value, it must be described in the coding file.</i>
*var11	ICDO-3 Topography	ICDO-3 definition with letter C	Use C999 if coded only in ICD-10	Please provide the full 4-character ICDO-3 code, but without the decimal point.  E.g. C504
*var12	ICDO-3 Morphology	ICDO-3 definition	Use 9999 if coded only in ICD-10	E.g. 8410
*var13	ICDO-3 Behavior	ICDO-3 definition	Not allowed	E.g. 3
*var14	ICD-10 Topography	ICD-10 definition	Use C999 if coded only in ICDO-3	E.g. C504
*var15	Basis of diagnosis	1-7	Not allowed	See definition of variables.
*var16	Clinical extent of disease	1=Localized 2=Regional 3=Distant metastases	9	See definition of variables.
*var17	TNM Staging		9	See definition of variables.
*var18	Last known vital status	1=Alive 2=Dead 3=Lost to follow-up	9	See definition of variables.
*var19	Date of vital status	YYYY-MM-DD	9999-99-99	At least year must be coded.  <i>If it is routine practice in your registry to substitute an imputed value for a missing value, it must be described in the coding file.</i>
*var20	Type of follow-up	1=Active 2=Passive	9	See definition of variables.
*var21	Method of follow-up	1-11	99	See definition of variables.

## 1.2. Definition of variables

### var4 Socioeconomic group (*optional*)

If the registry data permits analysis by socioeconomic group, the relevant category should be appropriately coded and *a description of the codes must also be submitted in the coding file*. Examples of socioeconomic group include education level, literacy status, employment level, household income, etc.

### var5 Ethnic group (*optional*)

If the registry data permits analysis by ethnic group, the relevant category should be appropriately coded and *a description of the codes must also be submitted in the coding file*.

### var8 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 requests of verification to your registry.

Code	Description
0	Failed
1	Tumor record has not been checked with IARC CHECK
2	Tumor record has been checked with IARC CHECK; no error(s) or warning(s)
3	Tumor record has been checked with IARC CHECK; any error(s) or warning(s) has been corrected
4	Tumor record has been checked with IARC CHECK; no change was made because the registry has confirmed that the original record was correct
9	This variable will not be provided

### var15 Basis of diagnosis

This variable indicates the degree of certainty with which a diagnosis of cancer has been established, in the specific context of survival analyses. The IARC recommends the following coding:

Code	Description	Criteria
0	Death certificate <b>only</b>	Information provided is from a death certificate.
<b>Non-microscopic</b>		
1	Clinical	Diagnosis made before death, but without any of the following (codes 2-7).
2	Clinical investigation	All diagnostic techniques, including X-ray, endoscopy, imaging, ultrasound, exploratory surgery (such as laparotomy), and autopsy, without a tissue diagnosis.
3	Specific tumor markers	Including biochemical and/or immunologic markers that are specific for a tumor site.
<b>Microscopic</b>		
5	Cytology	Examination of cells from a primary or secondary site, including fluids aspirated by endoscopy or needle; also includes the microscopic examination of peripheral blood and bone marrow aspirates.
6	Histology of a metastasis	Histologic examination of tissue from a metastasis, including autopsy specimens.

Code	Description	Criteria
7	Histology of a primary tumor	Histologic examination of tissue from primary tumor, however obtained, including all cutting techniques and bone marrow biopsies; also includes autopsy specimens of primary tumor.
9	Unknown	

**var16 Clinical extent of disease**

The spread of cancer as categorized by SEER summary staging version 2000 may be adopted as it is the basic categorization of stage. The clinical extent of disease before treatment has to be documented utilizing all the information from the medical records.

Code	Description
1	Localized
2	Regional
3	Distant metastases
9	Unknown if extension or metastasis (unstaged, unknown, or unspecified) Death certificate only case

**var17 TNM staging**

TNM staging defines the cancer spread based on the tumor size, to the extent that it has spread from the origin (primary site), and involvement of other organs in the body. The 7<sup>th</sup> edition of TNM classification manual published by the Union for International Cancer Control (UICC) will be followed.

Code	Description
<b>Primary Tumor (T)</b>	
TX	Primary tumor cannot be evaluated
T0	No evidence of primary tumor
Tis	Carcinoma in situ (CIS; abnormal cells are present but have not spread to neighboring tissue; although not cancer, CIS may become cancer and is sometimes called pre-invasive cancer)
T1, T2, T3, T4	Size and/or extent of the primary tumor
<b>Regional Lymph Nodes (N)</b>	
NX	Regional lymph nodes cannot be evaluated
N0	No regional lymph node involvement
N1, N2, N3	Degree of regional lymph node involvement (number and location of lymph nodes)
<b>Distant Metastasis (M)</b>	
MX	Distant metastasis cannot be evaluated
M1	Distant metastasis is present

**var18 Last known vital status**

The patient vital status known at the latest date of study closing date has to be recorded. The coding is:

Code	Description
1	Alive
2	Dead
3	Lost to follow-up
9	Vital status not known

Lost to follow-up could be coded when both active and passive method were adopted and the patient could not be followed up till the closing date or at 5 years of potential follow up. Unknown vital status would be coded when the patient could not be traced by both active and passive methods and the last known date is not later than the date of diagnosis.

### var19 Date of vital status

If the patient is dead, then the date of vital status is the date of death. If the vital status of the patient is alive, then the date corresponding to the most recent date of known vital status must be used.

### var20 Type of follow-up

Active follow-up refers to the efforts by the registry personnel in seeking information on the vital status of the. Passive follow-up is when mortality data is received by law from the vital statistics division or other agency for linking with the cancer incidence data and additional effort by the registry personnel is not required other than for the linking of the files.

Code	Description
1	Active
2	Passive
9	Unknown

### var21 Method of follow-up

Existing follow-up health information systems may be inadequate to provide complete follow-up data and may therefore lead to incomplete mortality ascertainment by passive method. Since survival analysis depends on the completeness of the mortality ascertainment, the registries must utilize efficient active method of follow-up if incomplete mortality ascertainment is expected. Some registries may also practice both the methods.

Code	Description
<b>Active</b>	
1	House visits/postal/telephone enquiry
2	Perusal of death certificate
3	Perusal of health or population registers maintained by official organization
4	Other reliable method (e.g. assessing case records at hospital)
<b>Passive</b>	
5	Electronic linkage with cancer mortality data with unique identification numbers
6	Electronic linkage with all-cause mortality data with unique identification numbers
7	Electronic linkage with cancer mortality data without unique identification numbers
8	Electronic linkage with all-cause mortality data without unique identification numbers
9	Electronic linkage with source hospital records
10	Linkage with population/public records
11	Receipt of notifications received from various sources (hospital/self)
<b>No information</b>	
99	Unknown

## 2. Population dataset

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 (2006-2012). 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.

**The population dataset, and the codes used in this file, should correspond to the incidence dataset with respect to time period, sex, age, socioeconomic group, ethnic group, and urban/rural setting.**

*Please provide references to all sources of population data submitted.*

<b>Table 3:</b> Description and format of all data variables for the population dataset (mandatory variables preceded by a *)				
Variable Name	Variable	Format	Unknown/Missing	Description
*var1	Year	YYYY	Not allowed	
*var2	Sex	1=Male 2=Female	Not allowed	
*var3	Age	Numeric	Not allowed	
*var4	Number of residents	Numeric	Not allowed	
var5	Socioeconomic group	Numeric	99	Same as incidence dataset
var6	Ethnic group	Numeric	99	Same as incidence dataset
var7	Urban or rural setting	1=Urban 2=Rural	9	

### File format

Each line of the population dataset 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 (e.g. 1=0-4 years of age, 2=5-9 years of age, ..., 17=80-84 years of age, and 18=85+ years of age). 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.

### 3. Mortality dataset (if available)

The mortality data should consist of all certified deaths from all causes among residents of the registration area during the same period as that covered by the incidence data (2006-2012). Wherever possible, the mortality data should be the official mortality data, as obtained from the Vital Statistics Department or equivalent and based on certificates/death records.

*Please provide references to all sources of mortality data submitted.*

<b>Table 4:</b> Description and format of all data variables for the mortality dataset (mandatory variables preceded by a *)				
Variable Name	Variable	Format	Unknown/Missing	Description
*var1	Year	YYYY	Not allowed	
*var2	Sex	1=Male 2=Female	Not allowed	
*var3	Age	Numeric	Not allowed	

**Table 4:** Description and format of all data variables for the mortality dataset (mandatory variables preceded by a \*)

Variable Name	Variable	Format	Unknown/ Missing	Description
*var4	<b>Number of deaths</b>	Numeric	Not allowed	
var5	Socioeconomic group	Numeric	99	Same as incidence dataset
var6	Ethnic group	Numeric	99	Same as incidence dataset
var7	Urban or rural setting	1=Urban 2=Rural	9	

\*\*please provide ICD edition if ICD-10 is not being used

### File Format

Each line of the mortality dataset should include number of deaths 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 (e.g. 1=0-4 years of age, 2=5-9 years of age, ..., 17=80-84 years of age, and 18=85+ years of age). 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. The total number of deaths can be provided if no breakdown information by age group is available.

### 4. Population life table dataset (if available)

The life table data should consist of all-cause mortality probabilities among residents of the registration area during the same period as that covered by the incidence data (2006-2012). The underlying mortality data should be obtained from the Vital Statistics Department or equivalent and based on certificates/death records.

*Please provide references to all sources for the life table data submitted.*

**Table 5:** Description and format of all data variables for the population life table dataset (mandatory variables preceded by a \*)

Variable Name	Variable	Format	Unknown/ Missing	Description
*var1	<b>Year</b>	YYYY	Not allowed	Needs to be provided for all incidence and follow-up years
*var2	<b>Sex</b>	1=Male 2=Female	Not allowed	
*var3	<b>Age</b>	Numeric	Not allowed	
*var4	<b>Mortality probability</b>	Numeric, 6 decimal places	Not allowed	

### File Format

Each line of the population life table dataset should include mortality probabilities for a combination of calendar year, sex and age. Age should be provided as a single year. Please provide the numbers of persons of unknown age, if applicable.

*If available, additional lifetables should be provided by socioeconomic group, ethnic group, and urban/rural setting using the same coding as in the incidence dataset.*

## 5. 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 include the following:

- Description of registry practice
- 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 socioeconomic and/or ethnic groups
- Any other information useful for processing the submitted dataset
- Any other information useful to evaluation of the results

### Downloads/Links

- Publications

SURVCAN-1: [View/Download](#)

SURVCAN-2: [View/Download](#)

- Software

IARCcrgtools: [Download](#)

- Websites

International Association of Cancer Registries (IACR): [Access](#)

International Agency for Research on Cancer (IARC): [Access](#)

Global Initiative for Cancer Registry Development (GICR): [Access](#)

Registries' Portal: [Access](#)

- Classification and Coding

International Classification of Diseases for Oncology (ICD-O-3): [Access](#)

European Network of Cancer Registries (ENCR) Recommendations for Coding of Incidence Data: [View/Download](#)

SEER Program Coding and Staging Manual: [View/Download](#)

IACR/IARC Multiple Primary Rules: [View/Download](#)

SEER Multiple Primary Rules: [View/Download](#)

Rules for coding of "Basis of Diagnosis" (IARC): [View/Download](#)