ENCR
European Network
of Cancer Registries

ENCR-JRC Call for Data, 2015

April 2015

European Commission | Joint Research Centre

ENCR is hosted by the European Commission's Joint Research Centre
Via E. Fermi 2749 | 21027 Ispra | Italy | jrc-encr@ec.europa.eu | www.encr.eu

TABLE OF CONTENTS

1. **INTRODUCTION**

The European Network of Cancer Registries (ENCR) began in 1989, within the framework of the Europe Against Cancer Programme of the European Commission, on the initiative of the International Agency for Research on Cancer (IARC), Association of Nordic Cancer Registries (ANCR), International Association of Cancer Registries (IACR) and Latin Language Registry Group (GRELL).

The ENCR is a professional, non-profit society dedicated to promoting collaboration between cancer registries (CRs), defining data collection standards and providing training for CR personnel.

It also aims to strengthen the basis for monitoring the cancer burden (incidence, mortality, prevalence and survival) in the European Union (EU), and Europe in a wider sense, through the provision of regular and timely information from European CRs.

The Joint Research Centre (JRC) is part of the European Commission. It consists of seven scientific Institutes located at five sites in five different Member States, with an administrative headquarters in Brussels. The JRC is the European Commission's in-house science service, which employs scientists to carry out research, and its mission is to provide independent scientific advice and support EU policy.

The Institute for Health and Consumer Protection (JRC-IHCP) is one of JRC's seven institutes and is based in Ispra (VA), Italy. JRC-IHCP provides scientific support to policy dossiers under the direction of the European Commission's Directorate-General for Health and Food Safety (SANTE).

Since 2012, JRC has hosted the ENCR secretariat and is active in supporting all ENCR activities, such as organising the regular ENCR Steering Committee meetings and in supporting cancer registration workshops and ENCR working groups. One of these working groups recently drafted a harmonised ENCR protocol for data-cleaning checks. This was approved by the ENCR and other cancer-registry stakeholders including all major European projects using CR data.

It is the wish of the ENCR to reduce the number of data calls to CRs as there is a heavy workload on CRs responsible for submitting cleaned data to various European studies. As a result it was agreed by the ENCR that there would be a single ENCR-JRC call for data, to serve all European projects. The protocol for this call serves the needs of all CR data users and avoids multiple separate calls foreseen in 2015 for studies such as EUROCARE or CONCORD.

JRC has, therefore, established a data portal on behalf of the ENCR, through which registries will be asked to submit their data for the 2015 call for data. Given the short time-frame for accomplishing this task, (due to the EUROCARE study issuing a call for data in early 2015), the ENCR-JRC portal will initially offer only essential functionality for data submission. It will, however, be rapidly upgraded to provide a comprehensive set of tools to best meet the needs of the registries.

In addition to the implementation of a data-submission portal, JRC and ENCR are developing data quality-checking (QC) software based on the data quality-check protocol[1] developed by the ENCR Quality Checks Working Group. JRC is currently devoting major resources (personnel and funds) to developing and providing a user friendly data-checking and quality control tool to the CRs, for the upcoming call. The QC software will be made freely available – also as source code – as soon as it has been fully developed and adequately tested.

Output from the European cancer incidence and mortality database will include a website showing basic information on temporal and geographical trends in cancer, by registry, and factsheets on individual cancers. It is also hoped to involve all participating stakeholders in contributing to peer-reviewed papers based on the aggregate data shown on the website. Individual data will also be made available to registries, researchers associated with them, and projects such as, for example, EUROCARE[2], Concord[3], Rarecare[4], and HAEMACARE, subject to specific permission from the registries. For these projects, permission can be given at the time of data submission.

2. REQUIRED DATA FILES

The following files should be submitted:

2.1 Cancer case file

2. 2 Population file

2.3 Mortality file

2.4 Life tables

2.5 Questionnaire

2.1 Cancer case file

Inclusion criteria for reportable incident cases:

Time period

All available registration years which are considered complete, up to and including, 2013.

Tumour

- All primary malignant tumours (behaviour=3), including basal cell and squamous cell carcinomas of skin.

  NOTE: data on non-melanoma skin cancers should be submitted only by registries which consider that registration of these cancers is essentially complete.
- All in situ tumours.
- All uncertain behaviour tumours.
- Benign tumours of the central nervous system (CNS) and urinary bladder

Multiple primary tumours

All multiple primary tumours are to be retained in the file. Multiple primary tumours will be defined according to the "International rules for multiple primary cancer (ICD-0 Third Edition)"[5].

Age

- All ages are eligible for registration.
- In age-restricted CRs, such as childhood CRs, all subsequent primary tumours of the registered patients should be included, if available, irrespective of age at diagnosis.

File format

The file should be formatted as follows:

- One record per tumour.
- The file should be a text file with semi colon (;) separators and should include a header, with variables named and order as specified in the text below (including the number and the text) and in Table 1.

  1_Flag; 2_Patient_ID; 3_Tumour_ID; 4_Day_DoB; 5_Month_DoB; 6_Year_DoB; 7_Sex; 8_Day_DoI; 9_Month_DoI; 10_Year_DoI; 11_Age; 12_BoD; 13_Topo; 14_Morpho; 15_Beh; 16_Grade; 17_Autopsy; 18_Vital_status; 19_Day_FU; 20_Month_FU; 21_Year_FU; 22_Survival; 23_Laterality; 24_Day_DoR; 25_Month_DoR; 26_Year_DoR; 27_Cause_death; 28_ICD_edition; 29_TNM_prefix; 30_pT; 31_pN; 32_pM; 33_cT; 34_cN; 35_cM; 36_Stage; 37_TNM_edition; 38_Cond_T; 39_Cond_N; 40_Cond_M; 41_Dukes; 42_FIGO; 43_AArbor; 44_Gleason; 45_Breslow; 46_EoD; 47_Tsize; 48_N_exam_nodes; 49_N_met_nodes; 50_Sent_nodes; 51_Met_sent_nodes; 52_Cfactor; 53_Surgery; 54_Systemic_th; 55_Radiotherapy; 56_BMtransp.
- Core variables must be submitted, for all records, according to the codes detailed in Table 1. Blank codes are not accepted for core variables.
- Additional variables should be submitted for all records, if collected. If, in the earlier years, these are not available, the fields should be blank and reports as “unknown”. If the CR does not systematically collect all the additional variables or only some of them, blank fields should be shown.
- All core and additional variables must be submitted for all records, whether blank or not, in the order shown in Table 1.

Data quality

- Data should be verified and corrected before submission. If ENCR-JRC data quality check software is not available, registries should use the IARCcrgTool software, downloadable from http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=68&Itemid=445. Records which generate errors or warnings should be flagged (see below “Check flag”).

Requested variables and coding

The description, format and type of variables, with corresponding missing/unknown values and coding schema, are summarised in Table 1.

Description of variables

- Check flag (1_Flag) (core variable)

Numeric variable, one digit.

Definition: This variable should be used to flag records, which have already been checked using the IARCcrgTool software, once errors and warnings have been corrected or the data verified.

Coding: Missing values are not allowed. Code values are: 0 (the record has not been checked) and 1 (the record has been checked with the IARCcrgTool software and verified or, if necessary, corrected). If the record has been coded 1, for the check flag variable, no request to check this record will be sent by JRC to the registry. If the code is 0 and, after checking with the ENCR-JRC Quality Check team, in collaboration with EUROCARE, errors or warnings are found, a request to check this tumour record will be sent.

- Patient identification code (2_Patient_ID) (core variable)

Alphanumeric variable, maximum fifty characters.

Definition: The patient identification is a unique code used by the registry, or generated at the time of data submission, to refer to each registered cancer patient. For data protection reasons, it should not be the official personal identification number.

Coding: Missing values are not allowed.

This code will be used for identifying patients with more than one primary tumour and for sending queries and logs to the CRs during the data cleaning process. If CRs have restrictions on providing the patient identification number used internally at the registry, a unique code can be created for each patient included in their data files. CRs need to keep a record of the correspondence between the patient identification number used in the registry and the code provided in their data submission.

- Tumour identification (3_Tumour_ID) (core variable)

Alphanumeric variable, maximum fifty characters.

Definition: This variable is allocated by the registry. It allows the identification of two or more tumours for the same patient. This can be, but does not need to be, a sequence number.

Coding: Missing values are not allowed.

The combination of the patient identification variable and the tumour identification variable should be unique for each tumour.

Date of birth

This consists of three separate core variables: day, month and year of birth.

A full and accurate date of birth is important because it will be used to calculate the exact age at diagnosis. This is used to calculate age-specific and age-standardised incidence and survival estimates. It is also used to calculate the exact age at death, and thus to select the appropriate background death rates from the life tables for the computation of expected survival.

- Day of birth (4_Day_DoB) (core variable)

Numeric variable, maximum two digits.

Definition: Day of birth of the patient.

Coding: The range of valid values is 1-31. Missing values are not allowed. If the day of birth cannot be provided, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the day of birth for some tumours has already been imputed, please provide us with any rules used to impute the day in the questionnaire (question 1.13).

- Month of birth (5_Month_DoB) (core variable)

Numeric variable, maximum two digits.

Definition: Month of birth of the patient.

Coding: The range of valid values is 1-12. Missing values are not allowed. If the month of birth cannot be provided, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the month of birth for some tumours has already been imputed, please provide us with any rules used to impute the month in the questionnaire (questions 1.13).

- Year of birth (6_Year_DoB) (core variable)

Numeric variable, four digits.

Definition: Year of birth of the patient.

Coding: Provide the full 4-digit year (for example 1942, 2004). The year of birth should not be less than 1842 (>1842). Missing values are not allowed.

- Sex (7_Sex) (core variable)

Numeric variable, one digit.

Definition: This variable refers to the biological and physiological characteristics that define men and women.

Coding: It should be coded as 1 (male), 2 (female) or 3 (other). Please do not exclude any records from your data on the basis of this variable, even if the sex of the patient is not known. In this case the code should be 9 (unknown).

Date of incidence

This consists of three separate core variables: day, month and year of incidence. ENCR recommendations should be followed to determine the date of incidence[6].

A full and accurate date of incidence is important, for both incidence and survival analysis, and is the starting date for the calculation of survival.

- Day of incidence (8_Day_DoI) (core variable)

Numeric variable, maximum two digits.

Definition: Record the day of incidence according to the ENCR recommendations[6]. The exact day of incidence is required to enable analysis of survival in the shortest follow-up intervals (first three months after diagnosis) and to study outcome indicators such as early death.

Coding: The range of valid values is 1-31. If the day of incidence is not available, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the day of incidence for some tumours has already been imputed, please provide us with any rules used to impute the day in the questionnaire (question 1.15).

- Month of incidence (9_Month_DoI) (core variable)

Numeric variable, maximum two digits.

Definition: Month of incidence recoded according to the ENCR recommendations[6].

Coding: The range of valid values is 1-12. If the month of incidence cannot be provided, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the month of incidence for some tumours has already been imputed, please provide us with any rules used to impute the month in the questionnaire (questions 1.15).

- Year of incidence (10_Year_DoI) (core variable)

Numeric variable, four digits.

Definition: Year of incidence recorded according to the ENCR recommendations[6].

Coding: This is a core variable and missing values are not allowed. The range of allowed values is from 1941 to present.

- Age at diagnosis (11_Age) (conditional core variable)

Numeric variable, maximum three digits.

This variable is **required if complete date of birth and/or incidence are not available**.

Definition: Latest completed year of age at the time of diagnosis.

Coding: The range of allowed values is 0-120.

- Basis of diagnosis (12_BoD) (core variable)

Numeric variable, one digit.

Definition: This variable indicates the degree of certainty with which the diagnosis of cancer has been established. This information will be used for providing quality indicators such as % Death Certificate Only (DCO) or % of cases with histological verification of diagnosis.

Coding: This variable should be coded according to the ENCR recommendations[7] as:

0 → Death certificate only (DCO)

1→ Clinical

2→ Clinical investigation

4→ Specific tumour markers

5→ Cytology

6→ Histology of a metastasis

7→ Histology of a primary tumour

9→ Unknown

Note: Cases registered as DCO are cancers for which no information could be obtained other than a death certificate mentioning cancer. These cases are included in cancer incidence statistics for the year of death. Nevertheless, the true date of diagnosis and the duration of the survival are unknown and these data cannot normally be included in survival analyses.

- Topography (13_Topo) (core variable)

Alphanumeric variable, four characters.

Definition: This variable indicates the anatomic site of the primary tumours.

Coding: It should be coded according to the third revision of the International Classification of Diseases for Oncology (ICD-O-3)[8]. CRs who use other ICD versions should convert their codes to ICD-O-3 prior to submission with IARCcrgTool software or another appropriate software.

The full four-digit character ICD-O-3 code should be provided, including the initial letter, but without the decimal point (“.”). For example, supraglottis should be coded as C321.

When the primary site of the tumour is unknown, topography of the metastasis should not be given and the topography should be coded as C809.

- Morphology (14_Morpho) (core variable)

Numeric variable, four digits.

Definition: The morphology code records the type of cell that has become neoplastic, the specific histological term.

Coding: This should be coded according to ICD-O-3[8]. CRs who use other ICD versions should convert their codes to ICD-O-3 prior to submission with IARCcrgTool software or another appropriate software.

The accepted range of morphology codes is 8000-9989. Blank fields or 9999 codes are not allowed. Malignant tumour, NOS, should be coded as 8000, leukaemia, NOS should be coded as 9800 and malignant lymphoma, NOS as 9590.

- Behaviour (15_Beh) (core variable)

Numeric variable, one digit.

Definition: The behaviour code indicates whether a tumour is malignant, benign, in situ, or of uncertain behaviour.

Coding: It should be coded according to ICD-O-3[8] as follows:

0→ Benign neoplasm

1→ Neoplasm of uncertain or unknown behaviour

2→ In situ neoplasm

3→ Malignant neoplasm stated or presumed to be primary

- Grade (16_Grade) (core variable)

Numeric variable, one digit.

Definition: This variable describes, how much or how little, a tumour resembles the normal tissues from which it arose and is also used to denote cell lineage for leukaemias and lymphomas.

Coding: It should be coded according to the ICD-O-3[8] as follow:

1→ Well differentiated

2→ Moderately differentiated

3 → Poorly differentiated

4→Undifferentiated, anaplastic

For leukaemias and lymphomas:

5→ T-cell; T-precursor

6→ B-cell; Pre-B; B-precursor

7→ Null cell; Non T-non B

8→ NK cell (natural killer cell)

For all:

9 → Unknown

When a diagnosis indicates two different degrees of grading or differentiation, the higher number should be used as the grade code. For example: if the diagnosis is moderately differentiated squamous cell carcinoma with poorly differentiated areas, the grade should be coded as 3.

- Incidental finding of cancer at autopsy (17_Autopsy) (core variable for survival studies)

Numeric variable, one digit.

Definition: Cases discovered only at autopsy. These cases are included in cancer incidence statistics. For these cases, the date of incidence is the same as the date of death. This variable is extremely important in survival analysis because cases incidentally discovered at autopsy, as well as DCO cases, must be excluded from survival statistics. For data to be included in survival studies this variable must be coded.

Coding: It should be coded as:

0 →No (not found at autopsy)

1 →Yes (found at autopsy)

9→Unknown

- Last known vital status (18_Vital_status) (core variable for survival studies)

Numeric variable, one digit.

Definition: This variable describes the patient's vital status as last known to the CR. This information may be collected using either 'active' or 'passive' methods of follow up.

Coding: It should be coded as:

1→ Alive

2→ Dead

9 →Unknown

If the CR uses passive follow-up, patients who are not known to be deceased would normally be assumed to be alive on the date of the most recent linkage between the registry and a death index or other vital status records. The vital status of those patients should be coded as "1" (alive).

If some patients cannot be traced by any active follow-up procedure, their vital status may remain undetermined and should then be coded as "9" (unknown).

Date of last known vital status

This is the most recent date for which the patient's last known vital status was available. If the patient is deceased, the date of last known vital status should be the date of death. For registries using passive follow-up this is the most recent date for which death certificates have been linked to registrations, which must be clearly notified when data are submitted.

If the patient has emigrated or has been lost to follow-up, the last date at which he/she was known to be alive should be reported.

- Day of last known vital status (19_Day_FU) (core variable for survival studies)

Numeric variable, maximum two digits.

Definition: The day of the most recent date for which the patient's last known vital status was available.

Coding: The range of valid values is 1-31. When the day of the last known vital status is not available, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the day of the last known vital status has already been imputed, please provide us with any rules used to impute the day (questions 1.17).

- Month of last known vital status (20_Month_FU) (core variable for survival studies)

Numeric variable, maximum two digits.

Definition: The month of the most recent date for which the patient's last known vital status was available.

Coding: The range of valid values is 1-12. When the month of the last known vital status cannot be provided, it should be coded as 99.

It would be preferable to have the data without imputation of the missing values. If, however, the month of the last known vital status for some tumours has already been imputed, please provide us with any rules used to impute the month (questions 1.17).

- Year of the last known vital status (21_Year_FU) (core variable for survival studies)

Numeric variable, four digits.

Definition: Year of the most recent date for which the patient's last known vital status was available.

Coding: The range of valid values is from 1941 to the present.

- **Duration of survival in days (22_Survival) (conditional core variable for survival studies)**

Numeric variable, maximum five digits.

This variable is required if complete date of incidence and/or date of last known vital status cannot be provided.

Definition: It is the number of days between the date of incidence and the date of the last known vital status.

Coding: The values should be ≥ 0. When the duration of survival cannot be provided, it should be coded as 99999.

- Laterality of paired organs (23_Laterality) (additional variable)

Numeric variable, one digit.

Definition: The paired organs for which such information may be relevant are: Parotid gland (C07), tonsil (C09), nasal cavity (C300), lung (C34), pleura (C384), long bones of upper limb and scapula (C400), short bones of upper limb (C401), long bones of lower limb (C402), short bones of lower limb (C403), rib and clavicle (C413), pelvic bones (excluding sacrum, coccyx, and symphysis pubis) (C414), skin of eyelid (C441), skin of external ear (C442), skin of arm and shoulder (C446), skin of leg and hip (C447), breast (C50), ovary (C56), fallopian tube (C570), testis (C62), epididymis (C630), kidney (C649), renal pelvis (C659), ureter (C66), eye (C69) and suprarenal gland (C74).

Coding: It should be coded as:

0→ Not applicable

1→ Right

2→ Left

3→ Unilateral NOS

4→ Bilateral

9→ Unknown

If the CR does not collect this information routinely, this variable can be left blank.

Date of case registration (additional variable)

This consists of three separate variables: day, month and year of case registration. This date is when the case was first recorded in the registry database. If the CR does not collect this date routinely, the day, month and year variables should be left blank.

- Day of case registration (24_Day_DoR) (additional variable)

Numeric variable, maximum two digits.

Definition: The day when the case was first recorded in the registry database.

Coding: If the CR does not collect this information routinely, this variable can be blank. When this information is available, the range of valid values is 1-31. When the day of case registration is not available for some cases, it should be coded as 99.

- Month of case registration (25_Month_DoR) (additional variable)

Numeric variable, maximum two digits.

Definition: The month when the case was first recorded in the registry database.

Coding: If the CR does not collect this information routinely, this variable can be blank. When this information is available, the range of valid values is 1-12. When the month of case registration is not available for some cases, it should be coded as 99.

- Year of case registration (26_Year_DoR) (additional variable)

Numeric variable, four digits.

Definition: The year when the case was first recorded in the registry database.

Coding: The range of allowed values is from 1941 to the present. When the year of case registration is not available for some cases, it should be coded as 9999.

If the CR does not collect this information routinely, this variable can be blank.

- **Official underlying cause of death (27_Cause_death) (additional variable)**

Alphanumeric variable, five characters.

Definition: This variable may be used to estimate cause-specific survival. It should be the official underlying cause of death according to standard international coding rules.

Coding: It should be coded according to the International Classification of Diseases (ICD). The dot (.) between the third and the fourth digits should not be included. For example, if the underlying cause of death is malignant neoplasm of laryngeal cartilage, this should be coded as C323 (according to ICD-10)[9]. If the underlying cause of death is acute myocardial infarction, unspecified, this should be coded as 4109 (according to ICD-9)[10].

If the CR does not collect this information routinely, this variable can be left blank.

- **ICD edition used for coding cause of death (28_ICD_edition) (additional variable)**

Numeric variable, maximum two digits.

This variable coded as 7, 8, 9 or 10 should be provided if the underlying cause of death has been reported.

Stage at diagnosis

Stage is particularly useful for the interpretation of international survival comparisons, for the evaluation of screening programs, and other studies. This information is often now available in satisfactory quality from many CRs.

Different coding schemes are used to classify tumour stage. Several optional variables have been provided to accommodate the different formats of data collected routinely by CRs.

When TNM stage and/or TNM stage grouping data are available, they should be reported in preference to any other coding system. The pathological stage should always be reported, if available. Clinical stage should be reported if pathological stage data are not available.

If the CR does not know if the TNM is pathological or clinical, this information should be included as clinical and be specified in the questionnaire.

When full TNM information is not complete, condensed TNM, as recommended by the ENCR Working Group on extent of disease[11], may be recorded. If neither TNM nor the condensed TNM are available, the summary extent of disease or one of the site-specific staging systems (e.g. Dukes, FIGO) may be used.

- TNM prefix (29_TNM_prefix) (additional variable)

Alphanumeric variable, one character.

Definition: This variable is for identification of special cases in TNM.

Coding: Two prefix modifiers will be considered:

y: stage assessed after neo-adjuvant therapy.

a: stage determined at autopsy.

If this variable is not applicable or not available, it should be left blank.

- **TNM stage, pathological primary site T (30_pT) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable encodes information on the extent of the primary tumour based on pathological evidence.

Coding: It should be coded according to any edition of the TNM classification, without "T". For example: **1a,** not T1a. When the information cannot be provided, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

- **TNM stage, pathological lymph nodes N (31_pN) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable provides information on the absence or presence and extent of regional lymph node metastasis, based on pathological evidence.

Coding: It should be coded according to any edition of the TNM classification, without "N". For example: **0**, not N0; **3a,** not N3a …

When this information is not available, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

- **TNM stage, pathological metastases M (32_pM) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable describes the absence or presence of distant metastasis, based on pathological evidence.

Coding: It should be coded according to any edition of the TNM classification without "M". For example **1a,** not M1a …

When this information is not available, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

- **TNM stage, clinical primary site T (33_cT) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable encodes information on the extent of the primary tumour, based on clinical evidence.

Coding: It should be coded according to the any edition of the TNM classification without "T". For example: **1a,** not T1a. When the information cannot be provided, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

- **TNM stage, clinical lymph nodes N (34_cN) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable provides information on the absence or presence and extent of the regional lymph node metastasis, based on clinical evidence.

Coding: It should be coded according to any edition of the TNM classification, without "N". For example: **0,** not N0; **3a**, not N3a …

When this information is not available, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

- **TNM stage, clinical metastases M (35_cM) (additional variable)**

Alphanumeric variable, maximum ten characters.

Definition: This variable describes the absence or presence of distant metastasis, based on clinical evidence.

Coding: It should be coded according to any edition of the TNM classification without M. For example: **0,** not M0; **1a**, not M1a …

When this information is not available, it should be coded as 9.

If the CR does not collect this information routinely, this variable can be left blank.

If the CR does not know if the TNM is pathological or clinical, this information should be included as clinical and be specified in the questionnaire (question 1.19.2).

TNM stage grouping (36_Stage) (additional variable)

Alphanumeric variable, maximum four characters.

Definition: This variable should preferably be based pathological TNM but, if pathological information is not available, it may be based on clinical TNM.

Coding: It should be coded according to any edition of the TNM classification. For example: I, IVC …

When this information is not available, it should be coded as 9. If the CR does not collect this information routinely, this variable can be left blank.

- **TNM edition (37_TNM_edition) (additional variable)**

Numeric variable, maximum two digits.

This variable should be provided when any TNM and /or TNM stage grouping have being reported. The allowed range of values is 5-7, or 99 when this information is not available.

If the CR does not collect staging routinely, this variable can be left blank.

Condensed TNM

When T, and/or N, and/or M codes have not been explicitly recorded in the clinical/pathological records, the CR should attempt to score the extent of disease according to the Condensed TNM following the ENCR recommendations[11].

- Condensed TNM T (38_Cond_T) (additional variable)

Alphanumeric variable, two characters.

Coding: This variable should be coded as follows, according to the ENCR recommendations[11]:

- TL→ Localised.

  This category comprises T1-2 tumours. Exceptions: T3 tumours of the thyroid, breast, melanoma (see below), and eye (except sarcoma of orbit) which are also to be considered localised.
- TA→ Advanced

  This category comprises T3-4 tumours. Exceptions: T2 tumours of ovary, fallopian tube, placenta, bone and soft tissues, which are to be considered advanced.
- TX→Unknown

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

Special condensed T codes for melanoma of the skin

- Code as localised (TL) those melanomas in TNM categories T1-3 (corresponding to Breslow thickness less than or equal to 4.00mm or Clark levels II-IV)
- Code as advanced (TA) those melanomas in TNM category T4 (corresponding to Breslow thickness greater than 4.00 mm or Clark level).

- **Condensed TNM, N (39_Cond_N) (additional variable)**

Alphanumeric variable, two characters.

Coding: This variable should be coded as follows, according to the ENCR recommendations[11]:

- N0→No regional nodes
- N1→Regional nodes
- NX→Unknown

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

- **Condensed TNM, M (40_Cond_M) (additional variable)**

Alphanumeric variable, two characters.

Coding: This variable should be coded as follows, according to the ENCR recommendations[8]:

- M0→No distant metastasis
- M1→Distant metastasis
- MX→Unknown

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

- **Dukes' stage 41_Dukes (additional variable).**

Alphanumeric variable, one character.

Definition: Dukes' stage[12] is a specialised classification of tumour stage for cancers of the colon and rectum only. The TNM classification is preferable because it is more detailed. Dukes' stage **should only be reported if TNM stage data are not available**.

Coding: This variable should be coded as follows:

- A → Dukes' stage A (this is equivalent to T1N0M0 or T2N0M0)
- B→ Dukes' stage B (this is equivalent to T3N0M0 or T4N0M0)
- C→ Dukes' stage C (this is equivalent to T(any)N1M0 or T(any)N2M0)
- D → Dukes' stage D (this is equivalent to T(any)N(any)M1)
- 8 → Not applicable
- 9 →Dukes' stage missing: no information on Dukes' stage

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

- **FIGO stage (42_FIGO) (additional variable)**

Alphanumeric variable, maximum five characters.

Definition: FIGO stage[13] is a specialised classification of tumour stage for cervical, ovarian and other gynaecological cancers. The TNM classification is preferable because it is more detailed. FIGO stage **should only be reported if TNM stage data are not available**.

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

Coding: Coding varies according to the tumour site. In the following, all possible codes were included:

- 0 → FIGO stage 0
- I → FIGO stage I
- IA→ FIGO stage IA
- IA1→ FIGO stage IA1
- IA2→ FIGO stage IA2
- IB → FIGO stage IB
- IB1 → FIGO stage IB1
- IB2 → FIGO stage IB2
- IC→ FIGO stage IC
- II →FIGO stage II
- IIA → FIGO stage IIA
- IIA1 → FIGO stage IIA1
- IIA2 → FIGO stage IIA2
- IIB → FIGO stage IIB
- IIC → FIGO stage IIC
- III → FIGO stage III
- IIIA → FIGO stage IIIA
- IIIB → FIGO stage IIIB
- IIIC → FIGO stage IIIC
- IIIC1 → FIGO stage IIIC1
- IIIC2 → FIGO stage IIIC2
- IVA → FIGO stage IVA
- IVB → FIGO stage IVB
- 8 → Not applicable
- 9 → FIGO stage missing: no information on FIGO stage

If the CR does not collect this information routinely, or has provided full TNM, it can be left blank.

- **ANN ARBOR stage (43_AArbor) (additional variable) (for Lymphomas only)**

Alphanumeric variable, maximum four characters.

Definition: ANN ARBOR stage[14] is the universally agreed method available for anatomic staging of Hodgkin and non-Hodgkin Lymphomas.

Since the TNM classification is not applicable for lymphomas, ANN ARBOR stage **should be reported**.

Coding:

- I → Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, or Waldeyer's ring)
- II → Involvement of two or more lymph node regions on the same side of the diaphragm (mediastinum is a single site; hilar lymph nodes are lateralized); the number of anatomic sites should be indicated by suffix (e.g., II3)
- III →Involvement of lymph node regions or structures on both sides of the diaphragm
  - III_1 With or without splenic, hilar, celiac or portal hepatic nodes
  - III_2 With para aortic, iliac or mesenteric nodes
- IV → Involvement of extranodal site(s) beyond these designated E (extranodal)
- 8 →Not applicable
- 9→Ann Arbor stage missing: no information on Ann Arbor stage

No information on the presence or absence of symptoms (Fever >38^{o} C, drenching sweats, weight loss [10% body weight over six months] is requested; No information on involvement of a single, extranodal site contiguous or proximal to known nodal site

- **GLEASON grading (44_Gleason) (additional variable) (for prostate cancer only)**

Definition: Gleason[15, 16] described five distinct patterns of prostate cancer growth which were based on the pattern that the tumour glands made (as viewed through the microscope on stained tissue sections). These are commonly referred to now as Gleason patterns 1 through 5, with pattern 1 being the best differentiated and, therefore, the most favourable. Most, if not all tumours, have a mixture of patterns.

To obtain a Gleason score or grade, the dominant pattern is added to the second most prevalent pattern to obtain a number between 2 and 10.

Report the highest Gleason score from the biopsy/TURP or prostatectomy/autopsy. Exclude results from tests performed after neoadjuvant therapy began. If the Gleason score is NOT available and the Gleason pattern is available, sum the two patterns up to obtain the Gleason score.

Coding:

Gleason Score at diagnosis (2 to 10)

1= Gleason ≤ 6

2=Gleason =7

3=Gleason 8-10

9=Gleason unknown: Gleason score cannot be processed

- **BRESLOW thickness (45_Breslow) (additional variable)**

Numeric variable, maximum three digits.

Definition: BRESLOW thickness[17] is one of the most important prognostic factors for skin melanoma. It measures the vertical thickness of the lesion in millimetres (mm).

Coding: It should be coded by numbers included in the range 0-900 mm. When this information is not available, it should be coded as 999.

If the CR does not collect this information routinely, this variable can be left blank.

- **Summary extent of disease (46_EoD) (additional variable)**

Numeric variable, one digit.

Definition: Summary stage is the most basic way of categorising how far a cancer has spread from its point of origin. Summary staging uses all the information available in the medical record; it is a combination of the most precise clinical and pathological evidence on the extent of disease. Many population-based CRs report summary stage for their registered cases because the staging categories are sufficiently broad to enable measurement of progress in cancer control.

Coding: This variable should be coded as follow:

1 → Tumour is confined to the topography of origin

2 → Tumour has spread to immediately adjacent tissues and/or regional lymph-nodes

3→ Tumour has spread to distant organs

4→ Tumour is not confined to the topography of origin but not specified whether code 2 or 3 applies

5→ No distant metastasis but not specified whether code 1 or 2 applies

9→ Unknown summary extent of disease

If the CR is not collecting this information routinely, it could be blank.

- **Tumour size in millimetres (mm) (47_Tsize) (additional variable)**

Numeric variable, maximum three digits.

Definition: This should represent the maximum tumour diameter in millimetres, and should be based on histological examination, if available.

Coding: This variable should be reported as numeric variable (millimetres). For example: 3, 10, etc.

Zero is not a valid tumour dimension. If no data are available, it should be coded as 999.

If the CR does not collect this information routinely, it can be left blank.

- **Number of nodes examined (48_N_exam_nodes) (additional variable)**

Numeric variable, maximum two digits.

Definition: This variable should report the exact number of lymph nodes examined, as recorded in the pathological records.

Coding: This variable should be coded as a numeric variable. If no information is available, or if pathological examination was not performed, it must be coded as 99.

If the CR does not collect this information routinely, it can be left blank.

- **Number of metastatic nodes (49_N_met_nodes) (additional variable)**

Numeric variable, maximum two digits.

Definition: This variable should report the exact number of lymph nodes containing tumour cells, as recorded in the pathological records.

Coding: This variable should be coded as a numeric variable. If no information is available, or if pathological examination was not performed, the code must be 99.

If the CR does not collect this information routinely, it can be left blank.

- **Sentinel nodes (50_Sent_nodes) (additional variable)**

Numeric variable, one digit.

Definition: This variable should report if sentinel nodes have been examined.

Coding: It should be coded as:

1→ Done (sentinel nodes have been examined)

2 → Not done (sentinel nodes have not been examined)

3 → Not applicable

9→ Unknown

If the CR does not collect this information routinely, it can be left blank.

- **Metastases in sentinel node (51_Met_sent_nodes) (additional variable)**

Numeric variable, one digit.

Definition: This variable should report if sentinel nodes contained tumour cells.

Coding: It should be coded as:

1 →Yes

2 → No

3 → Not applicable

9 → Unknown

If the CR does not collect this information routinely, it can be left blank.

- **C factor (52_Cfactor) (additional variable)**

Numeric variable, one digit.

Definition: This factor describes the likely validity of the stage data for a given case, in relation to the diagnostic methods used to determine it. It refers to the diagnostic examinations carried out to detect or exclude local extension and distant metastases.

Coding: The ENCR recommendations[11] should be used for coding this variable

1→ C1 Evidence from standard diagnostic methods only

2→ C2 Evidence obtained by special diagnostic means

3→ C3 Evidence from surgical exploration, including biopsy and cytology

4→ C4 Evidence following definitive surgery and pathological examination of the resected specimen

5→ C5 Evidence from autopsy

9→ Unknown

If the CR does not collect this information routinely, it can be left blank.

Treatment

Definition: The following variables are related to the first course of anticancer therapy after diagnosis. Purely symptomatic therapy (e.g. bypass surgery, pain relief) should not be included.

Coding: The treatment variables are additional, numeric and with one digit. They should be coded as follows:

- **Surgery (53_Surgery) (additional variable)**

Numeric variable, one digit.

1→ Yes (the patient has received this treatment)

2→ No (the patient has not received this treatment)

9→ Unknown

Surgery includes any operation to remove all or part of the cancer. Biopsy which is followed by definitive surgery should not be included; other biopsies, where the cancer is completely excised, can be included.

If the CR does not collect this information routinely, it can be left blank.

- **Systemic anti-cancer therapy, including chemotherapy, targeted therapy, immunotherapy and hormone therapy (54_Systemic_th) (additional variable)**

Numeric variable, one digit.

1→ Yes (the patient has received this treatment)

2→ No (the patient has not received this treatment)

9→ Unknown

If the CR does not collect this information routinely, it can be left blank.

- **Radiotherapy (55_Radiotherapy) (additional variable)**

  Numeric variable, one digit.

  1→ Yes (the patient has received this treatment)

  2→ No (the patient has not received this treatment)

  9→ Unknown

  If the CR does not collect this information routinely, it can be left blank.

- **Bone marrow transplantation (56_BMtransp) (additional variable)**

  Numeric variable, one digit.

  1→ Yes (the patient has received this treatment)

  2→ No (the patient has not received this treatment)

  9→ Unknown

  If the CR does not collect this information routinely, it can be left blank.

Table 1.Variable number, name, description, format, core/additional, missing/unknown values and coding schema

These variables should be separated by a semi-colon.

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Coding
1_Flag	Check flag	F	1	Y	Not allowed	0 → Not checked 1 →Checked
2_Patient_ID	Patient identification code	A	50	Y	Not allowed	According to registry coding
3_Tumour_ID	Tumour identification	A	50	Y	Not allowed	According to registry coding
4_Day_DoB	Day of birth	F	2	Y	99	Range of allowed values: From 1 to 31
5_Month_DoB	Month of birth	F	2	Y	99	Range of allowed values: From 1 to 12
6_Year_DoB	Year of birth	F	4	Y	9999	Range of allowed values: > 1842 and ≤ the current year
7_Sex	Sex	F	1	Y	9	1 →Male 2 → Female 3 →Other
8_Day_DoI	Day: date of incidence	F	2	Y	99	Range of allowed values: From 1 to 31
9_Month_DoI	Month: date of incidence	F	2	Y	99	Range of allowed values: From 1 to 12
10_Year_DoI	Year: date of incidence	F	4	Y	Not allowed	Range of allowed values: > 1941 and ≤ the current year
11_Age	Age at diagnosis (incidence date) in years	F	3	Y*	999	Range of allowed values: ≥ 0 and < 121
12_BoD	Basis of diagnosis	F	1	Y	9	0→Death certificate only (DCO) 1→Clinical 2→Clinical investigation 4→Specific tumour markers 5→Cytology 6→Histology of a metastasis 7→Histology of a primary tumour
13_Topo	Topography (topography of the metastasis is not admitted)	A	4	Y	Not allowed	Valid code in ICD-O-3
14_Morpho	Morphology	F	4	Y	Not allowed	Valid code in ICD-O-3
15_Beh	ICD-O-3 behaviour	F	1	Y	Not allowed	0→ Benign neoplasm 1→ Neoplasm of uncertain and unknown behaviour 2→ In-situ neoplasm 3→Malignant neoplasm stated or presumed to be primary

F: Numeric variable; A: Alphanumeric variable

Y= yes; N= no

*If complete date of birth and/or date of incidence are missing or unknown

Table 1 (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Coding
16_Grade	Grade(ICD-O-3)	F	1	Y	9	1→Well differentiated, 2→Moderately differentiated 3 →Poorly differentiated 4→Undifferentiated, anaplastic 5→T-cell; T-precursor 6→B-Cell; Pre-B; B-precursor 7→Null cell; Non T-non B 8→NK cell (natural killer cell)
17_Autopsy	Incidental finding of cancer at autopsy	F	1	Y	9	0→No 1→Yes
18_Vital_status	The last known vital status	F	1	Y	9	1→ Alive 2→ Dead
19_Day_FU	Day of last known vital status	F	2	Y	99	Range of allowed values: From 1 to 31
20_Month_FU	Month of last known vital status	F	2	Y	99	Range of allowed values: From 1 to 12
21_Year_FU	Year of last known vital status	F	4	Y	9999	Range of allowed values: > 1941 and ≤ the current year
22_Survival	Duration of survival in days	F	5	Y***	99999	≥ 0
23_Laterality	Laterality of paired organs	F	1	N	9	0→Not applicable 1→Right 2→Left 3→Unilateral NOS 4→Bilateral
24_Day_DoR	Day of case registration	F	2	N	99	Range of allowed values: From 1 to 31
25_Month_DoR	Month of case registration	F	2	N	99	Range of allowed values: From 1 to 12
26_Year_DoR	Year of case registration	F	4	N	9999	Range of allowed values: > 1941 and ≤ the current year
27_Cause_death	Official underlying cause of death	A	5	N	99999	International Classification of Diseases
28_ICD_edition	ICD edition used for coding cause of death	F	2	N	99	Range of allowed values: ≥ 7 and ≤ 10

F: Numeric variable; A: Alphanumeric variable
Y= yes; N= no
** If complete date of birth, data of incidence and/or date of end of follow-up are missing or unknown.
*** If complete date of incidence and/or date of end of follow-up are missing or unknown.

Table 1 (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Coding
29_TNM_prefix	Additional descriptor for TNM	A	1	N	Blank	Prefix modifiers will be considered: **y:** stage assessed after neo-adjuvant therapy; **a**: stage determined at autopsy
30_pT	TNM stage, pathological primary site (pT)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
31_pN	TNM stage, pathological lymph nodes (pN)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
32_pM	TNM stage, pathological metastases (pM)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
33_cT	TNM stage, clinical primary site (cT)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
34_cN	TNM stage, clinical lymph nodes (cN)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
35_cM	TNM stage, clinical metastases (cM)	A	10	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
36_Stage	TNM stage grouping	A	4	N	9	TNM Classification of Malignant Tumours, 5th, 6th or 7thedition
37_TNM_edition	TNM edition	F	2	N	99	Allowed values: 5, 6 and 7
38_Cond_T	Condensed TNM, T	A	2	N	9	TL→Localised TA→Advanced TX→Unknown
39_Cond_N	Condensed TNM, N	A	2	N	9	N0→No regional nodes N1→ Regional nodes NX→Unknown
40_Cond_M	Condensed TNM, M	A	2	N	9	M0→No distant metastasis M1→ Distant metastasis MX→Unknown
41_Dukes	Dukes' stage	A	1	N	9	A→Dukes' stage A B→Dukes' stage B C→Dukes' stage C D→Dukes' stage D 8→not applicable

F: Numeric variable; A: Alphanumeric variable
Y= yes; N= no

Table 1 (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing/ unknown values	Coding
42_FIGO	FIGO stage	A	5	N	9	0 →FIGO stage 0 I → FIGO stage I IA → FIGO stage IA IA1 → FIGO stage IA1 IA2 → FIGO stage IA2 IB → FIGO stage IB IB1 → FIGO stage IB1 IB2 → FIGO stage IB2 IC → FIGO stage IC II → FIGO stage II IIA → FIGO stage IIA IIA1 → FIGO stage IIA1 IIA2 → FIGO stage IIA2 IIB→ FIGO stage IIB IIB1→ FIGO stage IIB1 IIB2→ FIGO stage IIB2 IIC→ FIGO stage IIC III → FIGO stage III IIIA → FIGO stage IIIA IIIB→ FIGO stage IIIB IIIC→ FIGO stage IIIC IIIC1→ FIGO stage IIIC1 IIIC2→ FIGO stage IIIC2 IVA→ FIGO stage IVA IVB → FIGO stage IVB 8→not applicable
43_AArbor	ANN ARBOR stage	A	4	N	9	Allowed values: I,II,III,IV and suffixes 1,2 (for stage III) suffixes 2-9 (for stage II) 8→not applicable
44_Gleason	GLEASON grading	F	2	N	9	Range of valid values: 1 → Gleason ≤ 6 2→ Gleason =7 3 → Gleason 8-10
45_Breslow	BRESLOW thickness	F	3	N	999	Tumour size in mm. >0-900
46_EoD	Summary extent of disease	F	1	N	9	1→ Confined 2→Adjacent tissues, and/or regional lymph-nodes 3→Distant organs 4→Not confined but not specified whether code 2 or 3 applies 5→ Not distant metastasis but not specified whether code 1 or 2 applies
47_Tsize	Tumour size in mm	F	3	N	999	Range of allowed values: > 0 or 999

F: Numeric variable; A: Alphanumeric variable
Y= yes; N= no

Table 1 (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing/ unknown values	Coding
48_N_exam_nodes	Number examined nodes	F	2	N	99	Range of allowed values: From 0 to 99
49_N_met_nodes	Number metastatic nodes	F	2	N	99	Range of allowed values: From 0 to number examined nodes
50_Sent_nodes	Sentinel nodes	F	1	N	9	1 → Done 2 → Not done 3 → Not applicable
51_Met_sent_nodes	Metastatic in sentinel nodes	F	1	N	9	1 → Yes 2 → No 3 → Not applicable
52_Cfactor	C factor	F	1	N	9	1→C1 Evidence from standard diagnostic methods only 2→C2 Evidence obtained by special diagnostic means 3→C3 Evidence from surgical exploration, including biopsy and cytology 4→C4 Evidence following definitive surgery and pathological examination of the resected specimen 5→C5 Evidence from autopsy
53_Surgery	Surgery	F	1	N	9	1→ Yes 2→ No
54_Systemic_th	Systemic therapy	F	1	N	9	1→ Yes 2→ No
55_Radiotherapy	Radiotherapy	F	1	N	9	1→ Yes 2→ No
56_BMtransp	Bone marrow transplantation	F	1	N	9	1→ Yes 2→ No

F: Numeric variable; A: Alphanumeric variable
Y= yes; N= no

2.2 Population file

Information on population data should be provided from official censuses, from intercensal/postcensal estimates provided by Vital Statistics Departments, or equivalent, or other official sources. The population data should cover the same people as the cases (e.g. non-residents should be defined similarly).

Scope

The population data should correspond to the cancer case file with respect to:

- registration area
- time period, preferably for each individual calendar year.
- sex
- age-range

If possible, population figures should give the mid-year estimates for each sub-category.

File format

The file should be a text file with semi colon (;) separator and should include headers, with names as specified in examples 1 or 2.

The file should contain the following variables:

- Calendar year
- Sex
- Age: by single year of age, if possible, or otherwise 21 standard age-groups (see below)
- Number of residents

The variables in the population file should be in the same order as given above.

If the population data are available by single year of age and, for example, the period for the cancer cases is 1992-2013, the population file should be provided as in EXAMPLE 1.

If population data are not available by single year of age, 21 age groups should be provided, using the following codes: 1 (under 1 year), 2 (age group 1-4), 3 (age group 5-9), 4 (age group 10-14), 5 (age group 15-19),6 (age group 20-24), 7 (age group 25-29), 8 (age group 30-34), 9 (age group 35-39), 10 (age group 40-44), 11 (age group 45-49), 12 (age group 50-54), 13 (age group 55-59), 14 (age group 60-64), 15 (age group 65-69), 16 (age group 70-74), 17 (age group 75-79), 18 (age group 80-84), 19 (age group 85-89), 20 (age group 90-95), 21 (age group 95+). In this case, the population file should be provided as in EXAMPLE 2.

EXAMPLE 1

Calendar_ year	Sex	Age unit	Number of residents
1992	1	0	$N_{1992,1,0}$
1992	1	1	$N_{1992,1,1}$
1992	1	2	$N_{1992,1,2}$
1992	1	3	$N_{1992,1,3}$
1992	…	…	
1992	2	0	$N_{1992,2,0}$
1992	2	1	$N_{1992,2,1}$
1992	2	2	$N_{1992,2,2}$
1992	2	3	$N_{1992,2,3}$
…	…	…	
2013	1	100	$N_{2013,1,100}$
2013	2	0	$N_{2013,2,0}$
2013	…	…	
2013	…	100	$N_{2013,2,100}$

Sex = 1 → males; Sex=2 → females

EXAMPLE 2

Calendar_ year	Sex	Age group	Number of residents
1992	1	1	$N_{1992,1,1}$
1992	1	2	$N_{1992,1,2}$
1992	1	3	$N_{1992,1,3}$
1992	1	4	$N_{1992,1,4}$
1992	…		
1992	2	1	$N_{1992,2,1}$
1992	2	2	$N_{1992,2,2}$
1992	2	3	$N_{1992,2,3}$
1992	2	4	$N_{1992,2,4}$
…	…		
2013	1	21	$N_{2013,1,21}$
2013	2	1	$N_{2013,2,1}$
2013	…		
2013	…	21	$N_{2013,2,21}$

Sex = 1 → males; Sex=2 → females

Accompanying information required

- Any coding, other than that recommended, should be documented in the questionnaire.
- A reference to the source of population data should be provided in the questionnaire.

2.3 Mortality file

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.

These data will be used to calculate the ratio of mortality to incidence, by topography.

Scope

The mortality data for the area covered by a CR should include all residents whose underlying cause of death was cancer.

The mortality data should correspond to the cancer cases file with respect to:

- registration area
- time period, preferably for each individual calendar year
- sex
- age-range

File format

The file should be a text file with semi colon (;) separator and include headers, with reference to names as specified in examples 3 or 4.

The file should contain the following variables:

- Calendar year
- Sex
- Age: single year of age, if possible, or otherwise 21 age-groups (see below)
- Cause of death: 3 digits of the applicable International Classification of Diseases (ICD)
- Number of deaths

The variables in the mortality file should be in the same order as given above.

Mortality for all ages combined (total number of deaths) is acceptable only if no breakdown information by age-group is available to the registry.

If the number of deaths is available by single year of age and, for example, the period for cancer case file i.e. 1992-2013, the mortality file should be provided as in EXAMPLE 3.

Alternatively, the number of deaths for the combination of calendar year, sex, age group and cause of death should be provided (EXAMPLE 4), using the following codes: 1 (under 1 year), 2 (age group 1-4), 3 (age group 5-9), 4 (age group 10-14), 5 (age group 15-19), 6 (age group 20-24), 7 (age group 25-29), 8 (age group 30-34), 9 (age group 35-39), 10 (age group 40-44), 11 (age group 45-49), 12 (age group 50-54), 13 (age group 55-59), 14 (age group 60-64), 15 (age group 65-69), 16 (age group 70-74), 17 (age group 75-79), 18 (age group 80-84) and 19 (age group 85-89), 20 (age group 90-95), 21 (age group 95+).

EXAMPLE 3

Calendar year	Sex	Age unit	Cause of death	Number of Deaths
1992	1	0	C00	$N_{1992,1,0,C00}$
1992	1	1	C00	$N_{1992,1,1,C00}$
1992	1	2	C00	$N_{1992,1,2,C00}$
1992	1	3	C00	$N_{1992,1,3,C00}$
1992	…	…	…	…
1992	2	0	C00	$N_{1992,2,0,C00}$
1992	2	1	C00	$N_{1992,2,1,C00}$
1992	2	2	C00	$N_{1992,2,2,C00}$
1992	2	3	C00	$N_{1992,2,3,C00}$
…	…	…		
2013	1	100	C97	$N_{2013,1,100,C97}$
2013	2	0	C00	$N_{2013,2,0,C00}$
2013	…	…		
2013	…	100	C97	$N_{2013,2,100,C97}$

Sex = 1 → males; Sex=2 → females

EXAMPLE 4

Calendar year	Sex	Age group	Cause of death	Number of Deaths
1992	1	1	140	$N_{1992,1,1,140}$
1992	1	2	140	$N_{1992,1,2,140}$
1992	1	3	140	$N_{1992,1,3,140}$
1992	1	4	140	$N_{1992,1,4,140}$
1992	…			
1992	2	1	140	$N_{1992,2,1140}$
1992	2	2	140	$N_{1992,2,2,140}$
1992	2	3	140	$N_{1992,2,3,140}$
1992	2	4	140	$N_{1992,2,4,140}$
…	…			
2013	1	21	208	$N_{2013,1,21,208}$
2013	2	1	140	$N_{2013,2,1,140}$
2013	…			
2013	…	21	2008	$N_{2013,2,21,2008}$

Sex = 1 → males; Sex=2 → females

Accompanying information required

- Any coding, other than that recommended, should be documented in the questionnaire.
- A reference to the source of population data should be provided in the questionnaire.

2.4 Life tables

Life tables, i.e. the background mortality in the general population of the administrative territory covered by the cancer registry, must be provided by registries participating in survival studies.
Registries that participated in previous cycles of EUROCARE are requested to send an update (the most recent year) of follow up available (up to 2013). Registries participating in EUROCARE for the first time should send life tables covering their entire period of incidence and follow-up (up to 2013).

All-causes mortality rates in the general population, by sex, age and calendar year, should be provided to 6 decimal places or an equivalent number of significant figures (e.g. 0.012345 for a rate of 1,234.5 per 100,000). Since mortality rates are highly dependent on age, values should be preferably given by one-year age classes (from 0 to 99 or more). If this is not possible, age classes should be grouped by no more than five years: in this case, please specify how the life tables were smoothed.

It is essential to have accurate mortality data for the elderly to accurately estimate relative survival in this age group. The oldest age class can be open (e.g. 90 years and over), but the lower boundary of the oldest age class should not be less than 85 years.

For national cancer registries, life tables can be provided by administrative region (if there are significant differences in all causes mortality by region within the country). In this case the administrative region must also be provided for each patient record in the file of cancer cases.

Please document the source of demographic data and provide the name and e-mail address of a reference person to be contacted for more detailed information on demographic and mortality data in the general population, if further clarification is requested.

EXAMPLE 5

Calendar year	Sex	Annual age (years)	All causes mortality rate
1990	1	0	0.003228
1990	1	1	0.000272
1990	1	2	0.000376
1990	1	3	0.000379
..	…	…	…
1990	1	99	0.414117
1990	2	0	0.000379
1990	2	1	0.000376
1990	2	2	0.000373
…	…	…	
1990	2	99	0.389871
…	…	…	
2013	1	0	0.002528
2013	...	…	
2013	2	99	0.342862

Sex = 1 → males; Sex=2 → females

2.5 Questionnaire

The questionnaire is an essential part of the data submission process and for data interpretation and comparability among registries.

An invitation to fill in the questionnaire will be sent to the CRs immediately after data submission. The questionnaire is focused on the datasets submitted. EUSurvey, a European Commission tool, will be used for designing the questionnaire.

3. SUBMISSION INSTRUCTIONS

The following files should be submitted:

1. Cancer case file
2. Population file
3. Mortality file
4. Life-tables
5. Questionnaire

3.1 File preparation

There are no specific requirements for the names of the data files to be submitted in the ENCR-JRC portal. All data files should be provided as text files, with semi colon (;) separators and should include headers, with titles exactly as specified in the protocol.

The structure of the data files should be fixed according to the information provided in section 2. The variables in the data files should be in the order as given in section 2.

The questionnaire related to the data file submission will be available on the ENCR-JRC portal.

3.2 ENCR-JRC portal

CRs are required to submit their data files through the ENCR-JRC portal. Using the portal will allow the CRs to benefit from the following features:

- High security and speed of data transfer
- Automated data submission and its acknowledgement
- One submission for several European projects
- On-line confirmation of participation in the different European projects

Instructions for each step of the file uploading procedure are provided on the ENCR-JRC portal. Specific help will be available on request at JRC-ENCR@ec.europa.eu

Each CR will be able to access the ENCR-JRC portal with a specific username and password that will be communicated via email by the ENCR Secretariat.

Contacts

ENCR Secretariat

European Commission-Joint Research Centre

Institute for Health and Consumer Protection. Public Health Unit– Cancer Policy Support

Building 101, office 01/140

Via Enrico Fermi 2749

I-21027 Ispra (VA) - ITALY

Telephone: +39 0332 78 9926

E-mail: JRC-ENCR@ec.europa.eu

4. REFERENCES

1. Martos C, Crocetti E, Visser O, Rous B and the Cancer Data Quality Checks Working Group. A proposal on cancer data quality checks: one common procedure for European cancer registries. JRC Technical Reports, 2014.
2. European cancer registry based study on survival and care of cancer patients (EUROCARE). Available in: http://www.eurocare.it/.
3. Allemani C, Weir HK, Carreira H, et al, and the CONCORD Working Group. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2014; published online Nov 26. Available in: http://dx.doi.org/10.1016/S0140-6736(14)62038-9.
4. Information network on rare cancers (RARECARENET). Available in: http://www.rarecarenet.eu/rarecarenet/.
5. IARC/IACR/ENCR. International rules for multiple primary cancers (ICD-O Thirds Edition). IARC: Lyon, 2004. Available in http://www.encr.eu/images/docs/recommendations/MPrules_july2004.pdf
6. ENCR recommendations for coding Incidence. Available in: http://www.encr.eu/images/docs/recommendations/incideng.pdf
7. ENCR recommendation for coding Basis of Diagnosis. Available in: http://www.encr.eu/images/docs/recommendations/basisd.pdf
8. Percy C, Fritz A, Jack A, Shanmugarathan S, Sobin L, Parkin DM, Whelan S. International Classification of Diseases for Oncology (ICD-O). World Health Organization; 3 edition (December 2000). Available in: http://codes.iarc.fr/usingicdo.php
9. International Statistical Classification of Diseases and Related Health Problems, 10th Revision. World Health Organisation, 2011.
10. International Statistical Classification of Diseases and Related Health Problems, 9th Revision. World Health Organisation, 1977.
11. ENCR recommendation on Condensed TNM for Coding the Extent of Disease. Available in: http://www.encr.eu/images/docs/recommendations/extentofdisease.pdf
12. Dukes CE. The classification of cancer of the rectum. J Path Bacteriol. 1932; 35: 323-32.
13. Benedet JL, Bender H, Jones H, Ngan HY, Pecorelli S. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. Int J GynObs 2000; 70: 209-62.
14. Armitage JO. Staging Non-Hodgkin Lymphoma. CA: A Cancer Journal for Clinicians 2005; 55: 368–76.
15. National Comprehensive Cancer network Guidelines (version 1.2013) prostate cancers.

16. Adamo M, Dickie, L, Ruhl J. (January 2015). *SEER Program Coding and Staging Manual 2015.* National Cancer Institute, Bethesda, MD 20850-9765
17. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. Ann Surg 1970; 172: 902-8.