1 Introduction

1.1 CONCORD-2 will establish global surveillance of cancer survival. This annex to the main protocol provides a detailed description of the data that each registry is asked to provide. It has been substantially modified in the light of discussion at the CONCORD Working Group meeting in Cork, Ireland, 20-21 September 2012, attended by participants from 48 countries.

1.2 Data for the CONCORD study will be sent to the Cancer Survival Group at the London School of Hygiene and Tropical Medicine. We expect to receive up to 2,000 data files - 200 or more cancer registries may each supply up to 10 data files (one for each cancer). **To manage these files efficiently, we must impose strict rules on data structure.** This annex sets out the rules!

1.3 First, we define 10 **index sites** - the stomach, colon, rectum, liver, lung, breast (women), cervix, ovary and prostate, plus leukaemia.

1.4 Second, for simplicity, we will use "**cancer**" to refer to all invasive, primary, malignant neoplasms, including the haematological malignancies.

1.5 **Index cancers** are those that:
- occur at an **index site**
- were diagnosed in persons who are **normally resident** in the territory covered by the registry
- were diagnosed during the **calendar period** covered by the registry's data submission

1.6 Most participating registries will supply data on all 10 **index cancers**, but that is ***not*** a requirement in order to participate in CONCORD. Some specialised registries only record certain cancers (e.g., breast, colorectal, haematological; childhood cancers). Some registries do not have follow-up data for patients with every type of cancer that they register.

1.7 We will focus on analysing survival for **index cancers** diagnosed during the period **1 January 1995 to 31 December 2009**. You are invited to contribute data for as much of that period as possible, but it is ***not*** a requirement to supply data for the entire period 1995-2009. Some registries only began operation more recently than 1995.

1.8 When identifying the data you propose to submit, however, you should ensure that (a) the *incidence data* are considered to be complete for the years submitted and (b) the *follow-up* of all patients for their vital status is also considered to be complete, at least up to 31 December of the last year of incidence, and preferably a later year. **More recent data are admissible**, if your data are complete for 2010 or a later year. Some examples follow:
- Example (1): incidence data 1995-2009, follow-up to 31 December 2009
- Example (2): incidence data 1998-2007, follow-up to 31 December 2008
- Example (3): incidence data 2000-2010, follow-up to 31 December 2011

1.9 Some participating **cancer registries** will cover the entire national population; others will only cover part of a country (state, province, region, etc.).

1.10 Survival analyses will focus on **adults** (defined as aged 15-99 years at diagnosis). We will also examine survival from acute lymphoid leukaemia in **children** (0-14 years).

1.11 If your registry collects data on tumours that are benign (behaviour code 0), uncertain (1) or *in situ* (2), such as *in situ* cancers of the cervix, please include records for ***all*** neoplasms diagnosed at an **index site** when submitting your data. This will enable comparison of the intensity of diagnostic activity between participating registries (see variable 22, behaviour, in

Section 3). Please note, however, that **survival analyses will be restricted to invasive, primary malignancies (behaviour code 3).**

1.12 During the period 1995-2009, most registries switched to ICD-O-3[1] for coding tumour site, morphology and behaviour, instead of ICD-9[2], ICD-10[3] or ICD-O-2[4]. At the CONCORD Working Group meeting in Cork in September 2012, it was agreed to use ICD-O-3. If your data are not coded to ICD-O-3, ***please discuss this with us*** before submitting your data.

1.13 Full dates (day, month and year) of birth, diagnosis and death are important in international survival analyses. The evidence and the rationale are explained in a recent paper[5]. This is also available on the CONCORD web-site. A brief summary is given with the relevant variables in Section 3 of this document. If you *cannot* send the full dates of birth, diagnosis and death, ***please discuss this with us*** before preparing your data for submission. We may be able to help you obtain ethical approval, or to find an alternative solution.

1.14 **All cancer data files must have the same structure.** The details are set out for each variable in the following pages. First, please note the general points below, which apply to most variables:

- All data files will be tested for adherence to protocol. This is the first step of quality control. Tables of protocol adherence will be sent to you shortly after data submission. Data files that do not meet the protocol cannot be used. If you are in doubt about how to construct your data files, ***please discuss this with us*** before sending them.
- Every tumour record ***must contain a value for every variable***, both <u>core</u> and <u>optional</u> variables (see page 3):
  - *If you do not collect data* for a particular variable, you must still include that variable in every tumour record, whether it is a **core variable** or an **optional variable**. Do not leave the variable blank. For example, if you do not collect data on race/ethnicity (variable 8), you should assign the value 99 to that variable in every tumour record (see summary list on page 32).
  - *If no data are available* for a variable in a particular tumour record, you must still include that variable, whether it is a **core variable** or an **optional variable**. Assign the value 9, 99 or 999, etc., depending on how many digits are specified for that variable (see page 32). For example, if a tumour record contains no data for the core variable 'sex', you should assign the value 9 to that variable in every tumour record. Do not assign an imputed value.
  - *If you choose not to supply* an optional variable, you must still include that variable in every tumour record. Do not leave the variable blank. For example, if you choose not to supply the optional variables 23-37 on stage at diagnosis, you should assign the value 9, 99 or 999 (etc.) to these variables, depending on how many digits are specified for each variable (see page 32).
- If it is routine practice in your registry to substitute an imputed value for a missing value (e.g. the month of the year), and some of the variables you submit contain imputed values **please tell us**: *we will request a description* of the imputation procedures.
- If tumour records in your database include a special code ("flag") to indicate when a missing value has been imputed, ***please discuss this with us*** before submitting your data. *We will request a description* of how each flag has been generated.
- If you have modified a standard coding scheme (such as ICD-O-3) by adding special codes for local use in your registry, **please recode your data** to the standard form before submission. If you have any doubts about the appropriate conversion, ***please discuss this with us***.
- The variables on **stage at diagnosis** are optional, but many registries that do collect such data have requested survival analyses by stage. Stage data are only requested for cases **diagnosed in 2001 or later**. For cancers diagnosed during the period 1995-2000, therefore, the stage variables should be set to 9, 99, etc. (see page 32). Data on stage for 2000 or earlier years will not be included in quality control or survival analyses.

2 Variable names and short descriptions

2.1 The sequence of variables to be used in each record is as follows:

Core variables

Name	Short description
VAR1	Country
VAR2	Registry
VAR3	Person code
VAR4	Tumour code
VAR5	IARC check flag *
VAR6	Sex
VAR7	Region *
VAR8	Race/ethnicity *
VAR9	Day of birth
VAR10	Month of birth
VAR11	Year of birth
VAR12	Day of diagnosis
VAR13	Month of diagnosis
VAR14	Year of diagnosis
VAR15	Last known vital status
VAR16	Day of last known vital status
VAR17	Month of last known vital status
VAR18	Year of last known vital status
VAR19	Basis of diagnosis
VAR20	ICD-O-3 Topography
VAR21	ICD-O-3 Morphology
VAR22	Behaviour

Optional variables

Name	Short description
VAR23	SEER Summary Stage 2000
VAR24	Pathological T
VAR25	Pathological N
VAR26	Pathological M
VAR27	Clinical T
VAR28	Clinical N
VAR29	Clinical M
VAR30	Condensed T
VAR31	Condensed N
VAR32	Condensed M
VAR33	Dukes' stage
VAR34	FIGO stage
VAR35	Tumour size
VAR36	No. of lymph nodes examined
VAR37	No. of lymph nodes involved

* If your registry does not use these variables, see paragraph 1.14 on page 2

2.2 Details of the content and coding of each variable are given in Section 3.

2.3 Abbreviations

AJCC	American Joint Committee on Cancer
DCO	Death-certificate-only registration
FIGO	Fédération Internationale de Gynécologie et d'Obstétrique; International Federation of Gynecology and Obstetrics
IARC	International Agency for Research on Cancer (WHO)
ICCC	International Classification of Childhood Cancers
ICD	International Classification of Diseases (WHO)
ICD-O	International Classification of Diseases for Oncology (WHO)
LSHTM	London School of Hygiene and Tropical Medicine
NAACCR	North American Association of Central Cancer Registries
SEER	Surveillance, Epidemiology and End Results programme (US National Cancer Institute)
TNM	Tumour Nodes Metastasis (UICC)
UICC	Union for International Cancer Control

3 Description of variables

Variable 1 Country

Numeric variable, four digits.

The value for this variable is assigned centrally (see table below). It comprises a one-digit code for the continent followed by the 3-digit UN code for each country. The 4-digit code for your country in the table below *must be included in each tumour record*.

The names shown for each country in the table are mainly the English names associated with the UN code. We have shortened some of the names for convenience: this does not carry any political significance.

AFRICA	**Eastern Africa**		**North America**		**Northern Europe**
2404	Kenya	1124	Canada	4208	Denmark
2480	Mauritius	1840	United States of America	4233	Estonia
2638	Réunion			4246	Finland
2800	Uganda	**ASIA**	**Eastern Asia**	4352	Iceland
2716	Zimbabwe	3156	China	4372	Ireland
	Northern Africa	3344	China, Hong Kong SAR	4428	Latvia
2012	Algeria	3392	Japan	4440	Lithuania
2818	Egypt	3410	Korea	4578	Norway
2434	Libya	3158	Taiwan	4752	Sweden
2504	Morocco		**Southern Asia**	4826	United Kingdom
2788	Tunisia	3356	India		**Southern Europe**
	Southern Africa	3364	Iran	4191	Croatia
2710	South Africa	3586	Pakistan	4292	Gibraltar
	Western Africa		**South-Eastern Asia**	4380	Italy
2270	Gambia	3360	Indonesia	4470	Malta
2288	Ghana	3458	Malaysia	4620	Portugal
2466	Mali	3608	Philippines	4705	Slovenia
2566	Nigeria	3702	Singapore	4724	Spain
		3764	Thailand		**Western Europe**
AMERICAS	**Caribbean**		**Western Asia**	4040	Austria
5192	Cuba	3196	Cyprus	4056	Belgium
5312	Guadeloupe	3368	Iraq	4250	France
5474	Martinique	3376	Israel	4276	Germany
5630	Puerto Rico	3400	Jordan	4528	Netherlands
	Central America	3414	Kuwait	4756	Switzerland
5188	Costa Rica	3634	Qatar		
5484	Mexico	3792	Turkey	**OCEANIA**	**Australia and New Zealand**
	South America			9036	Australia
5032	Argentina	**EUROPE**	**Eastern Europe**	9554	New Zealand
5076	Brazil	4100	Bulgaria		**Melanesia**
5152	Chile	4203	Czech Republic	9540	New Caledonia
5170	Colombia	4348	Hungary		**Micronesia**
5218	Ecuador	4616	Poland	9316	Guam
5604	Peru	4642	Romania		**Polynesia**
5858	Uruguay	4643	Russian Federation	9258	French Polynesia
		4703	Slovakia		

If your country is not listed, please contact us for advice.

For reference, the country codes are at the following United Nations web-page, accessed 31 October 2012: http://unstats.un.org/unsd/methods/m49/m49regin.htm

Variable 2 Registry

Numeric variable; one to three digits (range 1-950).

We will provide the code for your registry before you prepare your data. The code for your registry must be included as variable 2 in every tumour record.

Together with the country code (variable 1), this variable will be used to link the data file with the relevant life tables during survival analysis.

Variable 3 Person code

Numeric variable, up to 15 digits; ***or***
Alphanumeric variable, up to 15 characters.

If the **person code** in your registry is numeric, you should not submit 'leading zeros'. For example, if the **person code** has nine digits, submit it as a nine-digit number (e.g. 123456789), and not as 000000123456789 (15 characters).

This is a unique code used in your **cancer registry** to refer to each registered cancer patient.

The **person code** can be any unique string of characters, but ***not*** the person's name, national identity number, social security number or any similarly recognisable code. The **person code** must be included in each tumour record, to enable you to check the record in the event that we identify possible errors during quality control.

The same **person code** must be included in any other tumour records supplied for the same person. Together with the **tumour code** (variable 4), this variable provides a unique identification of each tumour included in the study, for the purposes of quality control, but without compromising patient confidentiality.

Note: A few cancer registries do not routinely use a **person code**. These registries will need to create a unique code for each person included in their data files. The code will be used to identify patients with more than one **index cancer**. It will also enable the registry to identify all tumours for a given person in the event that we identify possible errors during quality control. If you have any doubts about the appropriate procedure, ***please discuss this with us***.

Variable 4 Tumour code

Numeric variable, up to ten digits, ***or***
Alphanumeric variable, up to ten characters.

If the tumour codes in your registry are numeric, you should not submit 'leading zeros'. For example, if your tumour code has six digits, submit it as a six-digit number (e.g. 123456), and not as 0000123456 (10 characters).

This is the code used in your **cancer registry** to refer to each registered tumour. Together with the **person code** (variable 3), this variable will enable persons with more than one **index cancer** to be identified.

The main survival analyses will include all primary, invasive, malignancies at an **index site** for patients diagnosed during the period 1995-2009 (or the calendar period for which your registry provides data – see paragraph 1.8 on page 1).

If your registry has submitted data for all patients diagnosed during 1995-2009, a patient with an invasive primary cancer of the breast diagnosed in 2000, followed by a different invasive primary cancer of the colon diagnosed in 2005, would therefore be included in the survival analyses for each of those cancers.

Variable 5 IARC CHECK flag

Numeric variable, one digit.

We will use this variable to avoid sending you requests to check tumour records that you have already checked and, if necessary, corrected (codes 2, 3 or 4).

Code		Meaning
1	=	Tumour record has not been checked with IARC CHECK
2	=	Tumour record has been checked with IARC CHECK: no error(s) or warning(s)
3	=	Tumour record has been checked with IARC CHECK: any error(s) or warning(s) have been corrected
4	=	Tumour 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

If you choose not to supply the IARC CHECK flag, ***please assign the code 9 to this variable*** in every tumour record.

Variable 6 Sex

Numeric variable, one digit.

Code		Meaning
1	=	Male
2	=	Female
9	=	Sex is ambiguous, or sex was not known

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.

Variable 7 Region

Numeric variable, up to five digits.

In some cases, it may be possible to estimate survival for geographic areas within the territory of your registry. For example, if your registry has national coverage, such analyses could be for regions (e.g. provinces, states, etc.) within your country. Alternatively, if your registry covers a province or state, such analyses could be for smaller regions (e.g. counties) within your territory.

The categories for geographic region will be different for each registry that supplies this variable.

If you wish us to provide such analyses, *you will need to include a suitable geographic code* in each tumour record.

You will also need to identify for us the region (province, state, county, etc.) that corresponds to each geographic code.

For example, the counties within some US states are a geographic variable of interest:

Code		**Meaning**
21001	=	Kentucky, Adair
21003	=	Kentucky, Allen
21005	=	Kentucky, Anderson
21007	=	Kentucky, Ballard
21009	=	Kentucky, Barren
21011	=	Kentucky, Bath
21...	=	Kentucky, ...

99999 = Region not known, or you will not supply this variable

We will also need to be able to construct appropriate life tables for each geographic region (province, state, etc.) for which you wish to obtain separate survival estimates: see Annex 6: Life tables).

If you choose not to supply data for separate areas of the territory covered by your registry ("Region"), ***please assign the code 99999 to this variable*** in every tumour record.

Variable 8 Race/ethnicity

Numeric variable, one or two digits.

In some cases, it may be possible to estimate survival separately for each race/ethnicity within a population. Cancer registries in some countries collect information on race and/or ethnicity (USA), race (Australia, Israel and Singapore), ethnicity (New Zealand, UK) or nationality (Dubai, Kuwait).

By contrast, most European registries ***do not*** record information on race or ethnicity. In some countries, it is illegal to do so.

The categories for race/ethnicity will be different for each registry that supplies this variable.

If you wish us to provide such analyses, *you will need to include a suitable code for race/ethnicity in each tumour record*.

You will also need to identify for us the race or ethnic group that corresponds to each code.

The example shown here is for the USA:

Code		Meaning
1	=	White, Hispanic
2	=	White, Non-Hispanic
3	=	White, Hispanic status unknown
4	=	Black, Hispanic
5	=	Black, Non-Hispanic
6	=	Black, Hispanic status unknown
7	=	Asian or Pacific Islander, Hispanic
8	=	Asian or Pacific Islander, Non-Hispanic
9	=	Asian or Pacific Islander, Hispanic status unknown
10	=	American Indian/Alaska Native, Hispanic
11	=	American Indian/Alaska Native, Non-Hispanic
12	=	American Indian/Alaska Native, Hispanic status unknown
13	=	Other, unspecified or unknown race, Hispanic
14	=	Other, unspecified or unknown race, Non-Hispanic
15	=	Other, unspecified or unknown race, Hispanic status unknown
99	=	Unknown or missing, or variable not supplied

Other race/ethnicity groups may be used, after discussion with the registry concerned, but the extent to which robust life tables can be constructed for each race or ethnic group may limit the scope for such analyses.

If you want us to provide survival analyses by race or ethnicity, we will need to construct appropriate life tables for each race/ethnicity: see Annex 6 (Life tables).

If your registry does not collect data on race/ethnicity, or you choose not to supply such data, ***please assign the code 99 to this variable*** in every tumour record.

Variables 9-11 Date of birth

A full and accurate date of ***birth*** is important because it is used to calculate the exact age at diagnosis. This is used to determine the age group (at diagnosis) into which patients are assigned for age-specific survival estimates, and later for age-standardised survival. It is also used to calculate the exact age at death, and thus to select the appropriate background death rate from the life table for the computation of expected survival.

A few cancer registries do not record the full date of birth. Most registries do record the full date, but some registries may face problems in supplying this information to external researchers, for a range of legal, ethical or regulatory reasons.

A brief explanation of why full dates are important is given in the **main protocol**. We have published a peer-reviewed article setting out the argument in detail, supported by empirical evidence.[5] The article shows the difficulties that arise in quality control when full dates cannot be obtained, and, more importantly, the biases that arise in survival estimation and survival comparisons. The conclusions are based on sensitivity analyses with a large data set. The article is accessible (with your login and password) on the CONCORD web-site.

Data preparation and analyses will be performed at the London School of Hygiene and Tropical Medicine (LSHTM). The Cancer Survival Group at LSHTM has acquired both statutory and ethical approvals from relevant bodies in the UK to receive and analyse individual tumour records with full dates (day, month and year) of birth, diagnosis and death for the CONCORD-2 study (**Annex 12.1: Statutory approval; Annex 12.2: Ethical approval**).

Variable 9 Day of birth

Numeric variable, one or two digits.

1-31 = the day of birth
99 = the day of birth of this patient is not known

Note: please tell us if the day *of the date of birth cannot be provided for* ***any*** *of your patients.*
Note: please see comments below *(variables 12-14) about the imputation of dates.*

Variable 10 Month of birth

Numeric variable, one or two digits.

1-12 = the month of birth
99 = the month of birth of this patient is not known

Note: please tell us if the month *of the date of birth cannot be provided for* ***any*** *of your patients.*
Note: please see comments below *(variables 12-14) about the imputation of dates.*

Variable 11 Year of birth

Numeric variable, four digits.

YYYY = the year of birth, from 1895 (person diagnosed in 1995 aged 99) to the present
9999 = the year of birth of this patient is not known

Variables 12-14 Date of diagnosis

The date of diagnosis should be the date used by the registry for cancer incidence or survival.

A full and accurate date of ***diagnosis*** is important because it is the starting point for the duration of survival.

A few cancer registries only record the month and year of diagnosis. Other registries only began to record the *full* date of diagnosis at some point since 1995.

A brief explanation of why full dates are important is given in the **main protocol**. We have published a peer-reviewed article setting out the argument in detail, supported by empirical evidence.[5] The article shows the difficulties that arise in quality control when full dates cannot be obtained, and, more importantly, the biases that arise in survival estimation and survival comparisons. The conclusions are based on sensitivity analyses with a large data set. The article is accessible (with your login and password) on the CONCORD web-site.

Some registries routinely capture more than one possible date of diagnosis (e.g. date of admission, date of biopsy, date of surgery, etc.). Before you supply your data for CONCORD, **you will need to complete the cancer registry questionnaire** on coding practices in your registry (**Annex 15**). This includes information about any rules that you use to select the date of diagnosis from two or more possible dates.

We kindly request that you do not impute the missing components of any dates while preparing your data for submission. If, however, the day and/or the month of the date of diagnosis for some tumours have *already* been imputed:

- ***Please provide us*** with any rules used to impute the day and/or the month of any dates
- If you routinely add a "flag" to tumour records to show when the day and/or the month of any date has been imputed, ***please discuss this with us*** before submitting your data.

Variable 12 Day of diagnosis

Numeric variable, one or two digits.

1-31 = the day of the date of diagnosis
99 = the day of the date of diagnosis of this patient is not known

Note: please tell us if the day *of the date of diagnosis cannot be provided for* ***any*** *of your patients.*

Variable 13 Month of diagnosis

Numeric variable, one or two digits.

1-12 = the month of the date of diagnosis
99 = the month of the date of diagnosis of this patient is not known

Note: please tell us if the month *of the date of diagnosis cannot be provided for* ***any*** *of your patients.*

Variable 14 Year of diagnosis

Numeric variable, four digits.

YYYY = the year of diagnosis, from 1995 onwards
9999 = the year of diagnosis of this patient is not known

Variable 15 Last known vital status

Numeric variable, one digit.

This variable encodes the patient's **last known vital status**, to the extent that it is known to the **cancer registry**. The ***date*** of the patient's **last known vital status** is recorded in variables 16-18.

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

Information about vital status is conventionally captured using either **'active'** or **'passive'** procedures, which we discuss below. Some registries use both. Before you submit a data file for CONCORD, **you will need to complete the cancer registry questionnaire (Annex 15)** on routine practices in your registry. If you have any doubts about which procedure is used for follow-up in your registry, ***please contact us before preparing your data***.

'Active' follow-up
Active follow-up refers to the situation in which the registry actively seeks information about the vital status of each cancer patient on a regular basis, e.g. from the patient's doctor, or hospital, or even home visits.

If your registry uses this approach, then you should use code "3" for patients whose vital status (alive or dead) could not be ascertained at the last vital status check: these patients are lost to follow-up. The last *date at which they were known to be alive* should be given in variables 16-18.

Code "3" should be used for patients known to have emigrated, since they are also lost to follow-up: the *date of emigration* should be given in variables 16-18.

'Passive' follow-up
Passive follow-up refers to the situation in which the registry routinely receives information from one or more reliable sources on the vital status of *all registered patients*. These sources vary widely between registries, but may include social security or health insurance files, or a regional or national index of persons who have died. The information may be derived in various ways, such as by computer linkage with the registry database, manual scanning of the death index, or supply of details about all deaths in the registry's territory.

The key features of passive follow-up for international survival comparisons are that:
- the registry uses this approach for updating its own data for local analyses of survival, and
- the registry's procedures reliably identify ***all deaths of registered cancer patients, not just the deaths for which cancer is mentioned on the death certificate***, and
- the registry can reliably assume that registered cancer patients are alive, unless information about a patient's death has been received from one or more of these sources.

If your registry uses passive follow-up, patients who are *not known to be dead* 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 would be coded as "1" (alive).

However, if some patients cannot be traced by any passive follow-up procedure, their vital status may remain undetermined: it would then be coded as "9" (unknown).

When you submit your data, *we will ask you to tell us* the last date when you performed linkage or other follow-up to determine the vital status of your patients.

Variables 16-18 Date of the patient's last known vital status

This is the most recent date for which the patient's **last known vital status** (variable 15) was available.

If the patient is dead, the **date of last known vital status** should be the date of death.

If the patient is *known* to be dead (variable 15 is coded as "2"), but the date of death is *not known*, the **date of last known vital status** should be coded to 99,99,9999 (see page 32).

If the patient has emigrated, the **date of last known vital status** should be the date of emigration.

If the patient has been lost to follow-up, the **date of last known vital status** should be the date of loss to follow-up.

If the patient is considered to be alive, but ***not*** emigrated or lost to follow-up, the **date of last known vital status** should be one of the following:

- 31 December of the last year for which follow-up of all patients is believed to be complete
- the date on which the registry last checked that patient's vital status, e.g. contact with the patient's doctor or a home visit (for registries that perform active follow-up), or linkage with a death index (for registries that perform passive follow-up) (see page 14)
- the date when the registry extracted the data file for this study from its database

Variable 16 Day of last known vital status

Numeric variable, one or two digits.

1-31 = the day of the date of last known vital status
99 = the day of this date is not known

Note: please tell us if the day of last known vital status cannot be provided for ***any*** of your patients.
Note: please see comments above (variables 12-14) about the imputation of dates.

Variable 17 Month of last known vital status

Numeric variable, one or two digits.

1-12 = the month of the date of last known vital status
99 = the month of this date is not known

Note: please tell us if the month *of last known vital status cannot be provided for* ***any*** *of your patients.*
Note: please see comments above *(variables 12-14) about the imputation of dates.*

Variable 18 Year of last known vital status

Numeric variable, four digits.

YYYY = the year of the date of last known vital status, from 1995 onwards
9999 = the year of this date is not known

Variable 19 Basis of diagnosis

Numeric variable, one digit.

This variable indicates the degree of certainty with which a diagnosis of cancer has been established, in the specific context of survival analysis:

Code Meaning

Not microscopically verified

1 = Clinical diagnosis ***only***
2 = Clinical investigation without a tissue diagnosis (e.g. endoscopy without biopsy, or imaging such as X-ray, ultrasound, computed tomography [CT] or magnetic resonance imaging [MRI])
3 = Clinical diagnosis, not otherwise specified [i.e. it is not known if code "1" or "2" applies]

Microscopically verified

4 = Cytologically confirmed (includes blood film examination for leukaemia)
5 = Histologically confirmed (includes bone marrow biopsy for leukaemia)
6 = Microscopically verified, not otherwise specified [i.e. not known if code "4" or "5" applies]

Evidence of cancer does not include the date of diagnosis

7 = Death-certificate-only registration (DCO) ***[see note]***
8 = Autopsy only - malignancy detected only at autopsy ***[see note]***

No information

9 = Unknown

Note:
Cancers registered ***solely*** on the basis of a death certificate (code 7) or autopsy (code 8) are usually included in cancer *incidence* statistics for the year in which they are registered.

For DCO and autopsy-detected cases, however, the true date of diagnosis and the duration of survival are unknown. Therefore, they cannot normally be included in *survival* analyses. A few cancer registries do not even record DCOs or autopsy-detected cancers.

If your cancer registry ***did*** register such cases during the calendar period covered by your data submission, however, ***you must include DCO and autopsy-detected cancers in the data you submit for this study***, to enable comparative quality control.

Variable 20 ICD-O-3 Topography

Alphanumeric variable, four characters.

Tumour site (topography) should be coded to the third edition of the International Classification of Diseases (ICD-O-3)[1].

Please provide the full 4-character ICD-O-3 code, but ***without the decimal point (".")*.** Thus, liver cancer will be C220 or C221 and prostate cancer will be C619. With this minor modification, the anatomic site of the **index cancers** will be coded as:

Stomach cancer: C160-C166; C168-C169

Colon cancer: C180-C189; C199

Rectal cancer: C209; C210-C212, C218
Note: includes anus and anal canal, C210-C218.

Liver cancer: C220-C221
Note: includes intrahepatic bile ducts, C221.

Lung cancer: C340-C343; C348-C349
Note: trachea (ICD-O-3 C339) will not be included in this study.

Breast cancer: C500-C506; C508-C509

Cervical cancer: C530-C531; C538-C539

Ovarian cancer: C480-C482; C569; C570-C574; C577-C579
Note: includes peritoneum and retroperitoneum, C480-C482, where cancers of high-grade serous morphology often originate in the fallopian tube, C570
Note: includes other and unspecified female genital organs, C577-C579.

Prostate cancer: C619

Leukaemia: You should select **leukaemias** for your data file **on the basis of their morphology code (variable 21), and *not*** on the basis of their topography code. You can use ***any*** valid ICD-O-3 topography code, but without the decimal point ("."), as specified above.

Variable 21 ICD-O-3 Morphology

Numeric variable, four digits.

Tumour morphology should be coded to the third edition of the International Classification of Diseases for Oncology (ICD-O-3)[1].

For microscopically confirmed tumours, the ICD-O-3 range of morphology codes is:

8000-9989

For haematological malignancies, the ICD-O-3 range of morphology codes is:

9590-9989

This range of morphology codes should be used to select all leukaemias and other haematological malignancies, including the lymphomas. This range of codes is the same for children and adults.

Leukaemia is the **only index cancer** for which you should use the ICD-O-3 morphology code to select cases for your data submission: all tumour records with an ICD-O-3 morphology code in the range 9590-9989 should be included. The **other nine index cancers** should be selected on the basis of the ICD-O-3 *topography* code (variable 20).

Selection and grouping of the adult leukaemias for survival analyses will be based on the categories established by a consensus of haematologists, pathologists and epidemiologists in the HAEMACARE Working Group[6-8]. The HAEMACARE manual for coding and reporting haematological malignancies exists in English[9] and Spanish (available on request).

Childhood leukaemias will be grouped differently from the adult leukaemias, on the basis of the third revision of the International Classification of Childhood Cancers[10] (ICCC-3). The ICCC-3 groupings are based on ICD-O-3 morphology codes.

For solid tumours *without* microscopic verification, you should use:

9999 This means the absence of data in the CONCORD study; it is not an ICD-O-3 code.

Note: this code is only valid for solid tumours, not leukaemias. Leukaemias are ***defined by their morphology***, so they must have a morphology code in the range 9590-9989.

Variable 22 Behaviour

Numeric variable, one digit.

Survival analyses will only include invasive, primary, malignant neoplasms. However, we will also report the *distribution of tumour behaviour* for each cancer. This will enable comparison of the intensity of diagnostic activity between contributing areas, e.g. the proportion of women with cervical cancer who were registered with *in situ* carcinoma.

Therefore, if your registry collects data on tumours that are benign (behaviour code 0), of uncertain behaviour (1) or *in situ* (2), such as *in situ* carcinoma of the cervix, ***please include records for all neoplasms*** (behaviour codes 0-3) diagnosed at an **index site** when submitting your data.

Please do *not* select tumours for inclusion in your data files on the basis of tumour behaviour.

Tumour behaviour should be coded to the third edition of the International Classification of Diseases for Oncology (ICD-O-3). The coding of tumour behaviour has been the same in all revisions of ICD-O.

Code		Meaning
0	=	Benign
1	=	Uncertain whether benign or malignant
2	=	Carcinoma *in situ*
3	=	Malignant, primary site

The behaviour codes below are included in ICD-O-3, but they are not usually used by cancer registries. We show them for completeness.

If your data ***do*** include behaviour codes 6 and 9, however, ***please do not recode them*** before data submission. Instead, ***please provide us with a description*** of how the codes have been used in your data:

Code		Meaning
6	=	Malignant, metastatic site
9	=	Malignant, uncertain whether primary or metastatic site

STAGE OF DISEASE AT DIAGNOSIS **Optional**

Provision of data on stage at diagnosis (variables 23-37) is *optional*.

"Tumour stage" describes how far the cancer has spread at the time of diagnosis. It is a key determinant of survival (prognostic factor). Knowledge of the stage at diagnosis of cancer patients is increasingly important for the interpretation of international survival comparisons[11-15]. Where possible, we will perform analyses of survival in relation to stage at diagnosis.

Among the registries that collect data on stage, information of satisfactory quality is often available for most of the 10 **index cancers**, at least for more recent years.

If your registry ***does not collect any data on tumour stage***, you should assign the codes 9, 99 or 999 (etc.) to ***all*** stage variables (variables 23-37) in every tumour record (see list on page 32). However, ***you must ensure that your data files meet the overall specification***, as summarised on page 32 of this Annex!

If your registry ***does collect data on stage***, but you choose ***not*** to submit data for one or more stage variables, please assign the codes 9, 99 or 999 (etc.) to those variables in every tumour record, depending on how many digits are specified for those variables (see list on page 32).

If you ***do*** submit data on stage, please supply stage data ***only for patients diagnosed from 1 January 2001 onwards***. For patients diagnosed before 2001, please assign the codes 9, 99 or 999 (etc.) to ***all*** stage variables (variables 23-37) in every tumour record, depending on how many digits are specified for those variables (see list on page 32). Data on stage at diagnosis for patients diagnosed before 2001 will *not be included* in quality control or survival analyses.

Many different coding schemes are being used to categorise tumour stage in registries around the world[16].

We will try to obtain data on **at least one widely used categorisation of stage** at diagnosis for each tumour (details in pages 21-31), to enable analysis of survival by stage at diagnosis:

SEER Summary Stage 2000
TNM (both pathological and clinical)
Condensed TNM
Dukes' stage (colon and rectum)
FIGO stage (cervix and ovary)
Tumour size
No. of lymph nodes examined
No. of lymph nodes positive for tumour

If you wish to supply data on stage at diagnosis, but your data on stage are not coded ***either*** to TNM 7th edition ***or*** to any of the other specific classifications in this list, ***please contact us before submitting your data.***

Variable 23 SEER Summary Stage 2000 Optional

Numeric variable, one digit.

SEER Summary Stage 2000 is a simple categorisation of stage, developed by the US National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) programme. The North American Association of Central Cancer Registries uses SEER Summary Stage 2000[17] (http://seer.cancer.gov/tools/ssm/). We borrow text from the introduction here. It has been in use in the US and Canada since 1 January 2001.

"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 for the extent of disease**. Many population-based cancer registries report summary stage for their registered cases, because the staging categories are sufficiently broad to enable measurement of progress in cancer control.

We expect that North American registries will supply SEER Summary Stage 2000 coded directly for cases diagnosed 2001-2003, but derived from Collaborative Stage for cases diagnosed in 2004 and later (http://seer.cancer.gov/tools/collabstaging). If your registry plans to supply SEER Summary Stage 2000 coded in any other way, ***we request that you inform us*** of the procedures you have used. The comparability of these staging schemes over time is addressed on the following SEER web-page:
http://seer.cancer.gov/seerstat/variables/seer/yr1973_2009/lrd_stage/index.html

Regional spread of disease is divided into several categories, according to the method of spread of the cancer:

Code		Meaning
0	=	*In situ*
1	=	Localised only
2	=	Regional spread by direct extension ***only***
3	=	Regional lymph nodes involved ***only***
4	=	Regional spread by ***both*** direct extension ***and*** lymph node involvement
5	=	Regional, NOS (not otherwise specified) – use this code if there is regional spread of the cancer, but ***the route of spread is not known***
7	=	Distant site(s) or lymph node(s) are involved
9	=	Unknown if there is extension or metastasis (unstaged, unknown or unspecified), ***or*** this is a ***death-certificate-only*** case, ***or*** this is an ***autopsy-only*** case

TNM stage (variables 24-29) **Optional**

The Tumour-Nodes-Metastasis (TNM) classification of stage at diagnosis uses a combination of clinical and pathological evidence, like SEER Summary Stage 2000.

The TNM classification is published by the Union for International Cancer Control (UICC). We will use the 7th edition of the TNM manual[18]. This is identical to the classification published by the American Joint Committee on Cancer (AJCC) in 2009[19]. If you have stage data that are coded to earlier editions of TNM, ***please contact us before submitting your data.***

The three components of TNM are tumour size (T); the status of regional lymph nodes, i.e. the extent of lymph node invasion by tumour (N), and whether there is metastasis (spread of disease to an organ or organs distant from the organ of origin) (M).

TNM stage data may be based on pathological evidence ("p") or clinical evidence ("c").

Variable 24 Pathological T **Optional**

Numeric variable, one digit.

This variable encodes information on the physical size of the tumour.

For the nine solid **index cancers** (i.e. excluding leukaemia), up to 4 sub-categories (a, b, c, d) exist for each of the stage categories pT1, pT2, pT3 and pT4. These sub-categories should be coded in the same way as the parent category: for example, pT1a should be coded to "1", in the same way as pT1.

For **cervical cancer only**, additional sub-categories exist: pT1a1, pT1a2, pT1b1 and pT1b2 should be coded to "1", in the same way as pT1. Similarly, sub-categories pT2a1 and pT2a2 should be coded to "2", in the same way as pT2.

The following codes will be used:

Code Meaning

0 = pT0 – no histological evidence of primary tumour

1 = pT1 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories pT1a, pT1b and pT1c, as well as for sub-categories pT1a1, pT1a2, pT1b1 and pT1b2 (cervix only)

2 = pT2 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories pT2a, pT2b and pT2c, as well as for sub-categories pT2a1, pT2a2 (cervix only)

3 = pT3 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories pT3a and pT3b

4 = pT4 – tumour of any size, with direct extension to adjacent organs
This code should also be used for sub-categories pT4a, pT4b, pT4c and pT4d

8 = is – *in situ* carcinoma

9 = pTX – unknown: primary tumour cannot be assessed histologically

Variable 25 Pathological N Optional

Numeric variable, one digit.

This variable encodes the extent of involvement of regional lymph nodes with tumour.

For cancers of the stomach, colon, rectum and breast, several subcategories of pN1, pN2 and pN3 exist (a, b and c). These should be coded in the same way as the parent category: for example, pN2b should be coded to “2”, in the same way as pN2.

The following codes will be used:

Code		Meaning
0	=	pN0 – no regional lymph nodes involved with tumour, histologically
1	=	pN1 – the content of this category varies with the cancer (see TNM manual[18]) *For colon, rectum and breast, this code should also be used for sub-categories pN1a, pN1b and pN1c*
2	=	pN2 – the content of this category varies with the cancer (see TNM manual[18]) *For colon, rectum and breast, this code should also be used for sub-categories pN2a, pN2b and pN2c*
3	=	pN3 – the content of this category varies with the cancer (see TNM manual[18]) *For stomach, this code should also be used for sub-categories pN3a and pN3b* *For breast, this code should also be used for sub-categories pN3a, pN3b and pN3c*
9	=	pNX – unknown: regional lymph nodes cannot be assessed histologically

Variable 26 Pathological M Optional

Numeric variable, one digit.

This variable encodes information on the presence or absence of distant metastases.

For cancers of the colon, rectum and prostate, several subcategories of pM1 exist (a, b and c). These should be coded to “1”, in the same way as for pM1.

The following codes will be used:

Code		Meaning
1	=	pM1 – Distant metastases have been microscopically confirmed *This code should also be used for sub-categories pM1a, pM1b and pM1c*
9	=	Unknown – this is not a TNM code (see below). It should be used when no data are available on pathological M status

Note:
The codes pM0 and pMX are ***not valid*** in TNM 7th edition[18]: see note on variable 29 below.

Variable 27-29 Clinical TNM stage **Optional**

Clinical data on tumour stage may be available from clinical examination, or from imaging of the tumour (X-ray, computed tomography [CT], magnetic resonance imaging [MRI], etc.).

Clinical data may be the only available data on tumour stage, if no surgical or invasive diagnostic procedure has been performed.

These variables are optional. If you choose not to supply them, please assign the code 9 to these variables in every tumour record.

Variable 27 Clinical T **Optional**

Numeric variable, one digit.

This variable encodes information on the physical size of the tumour.

Clinical data on the T component of stage should ***only*** be submitted if pathological data (variable 24) are ***not*** available.

For the nine solid **index cancers** (i.e. excluding leukaemia), up to 4 sub-categories (a, b, c, d) exist for each of the stage categories cT1, cT2, cT3 and cT4. These sub-categories should be coded in the same way as the parent category: for example, cT1a should be coded to "1", in the same way as cT1.

For **cervical cancer only**, additional sub-categories exist: cT1a1, cT1a2, cT1b1 and cT1b2 should all be coded to "1", in the same way as cT1. Similarly, sub-categories cT2a1 and cT2a2 should both be coded to "2", in the same way as cT2.

The following codes will be used:

Code Meaning

0 = cT0 – no evidence of primary tumour

1 = cT1 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories cT1a, cT1b and cT1c, as well as for sub-categories cT1a1, cT1a2, cT1b1 and cT1b2 (cervix only)

2 = cT2 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories cT2a, cT2b and cT2c, as well as for sub-categories cT2a1, cT2a2 (cervix only)

3 = cT3 – the content of this category varies with the cancer (see TNM manual[18])
This code should also be used for sub-categories cT3a and cT3b

4 = cT4 – tumour of any size, with direct extension to adjacent organs
This code should also be used for sub-categories cT4a, cT4b, cT4c and cT4d

8 = is – *in situ* carcinoma

9 = cTX – unknown: primary tumour cannot be assessed

Variable 28 Clinical N Optional

Numeric variable, one digit.

This variable encodes information on the involvement of regional lymph nodes with tumour.

Clinical data on the N component of stage should ***only*** be submitted if pathological data (variable 25) are ***not*** available.

For cancers of the stomach, colon, rectum and breast, several subcategories of cN1, cN2 and cN3 exist (a, b and c). These should be coded in the same way as the parent category: for example, cN2b should be coded to "2", in the same way as cN2.

The following codes will be used:

Code Meaning

0 = cN0 – no regional lymph nodes involved with tumour

1 = cN1 – the content of this category varies with the cancer (see TNM manual[18])
For colon, rectum and breast, this code should also be used for sub-categories cN1a, cN1b and cN1c

2 = cN2 – the content of this category varies with the cancer (see TNM manual[18])
For colon, rectum and breast, this code should also be used for sub-categories cN2a, cN2b and cN2c

3 = cN3 – the content of this category varies with the cancer (see TNM manual[18])
For stomach, this code should also be used for sub-categories cN3a and cN3b
For breast, this code should also be used for sub-categories cN3a, cN3b and cN3c

9 = cNX – unknown: regional lymph nodes cannot be assessed

Variable 29 Clinical M Optional

Numeric variable, one digit.

This variable encodes information on the presence or absence of distant metastases.

For cancers of the colon, rectum and prostate, several subcategories of cM1 exist (a, b and c). These sub-categories should be coded to “1”, in the same way as for cM1.

The following codes will be used:

Code		Meaning
0	=	cM0 – No metastases
1	=	cM1 – Metastases

Note:
If the clinician does not record the presence of metastases, ***it is assumed*** under the TNM 7th edition that no metastases are present (cM0): such cases should be coded to “0”.

The code “MX” was used in earlier editions of TNM to indicate that the metastatic status of the tumour was unknown. However, clinical assessment of metastasis can be based on physical examination alone, so cMX is no longer considered an appropriate code.

The code cMX is ***not valid*** in TNM 7th edition.

Variables 30-32 Condensed TNM **Optional**

The condensed TNM scheme for recording tumour stage was developed by the European Network of Cancer Registries[20] for tumour records in which the individual values of T and/or N and/or M are not explicitly recorded. Condensed TNM is based on the TNM 6th edition[21].

Condensed TNM data are only requested if ***neither*** pathological TNM data (variables 24-26) ***nor*** clinical TNM data (variables 27-29) are available.

There is a direct correspondence with the simplified stage classification that is often recorded by cancer registries, in which the extent of disease is classified as localised, regional or distant.

Variable 30 Condensed T **Optional**

Numeric variable, one digit.

Code	Meaning
1 =	L - Localised disease *Localised disease means:* T1 and T2 tumours for cancers of the stomach, colon, rectum, liver, lung, cervix, prostate T1, T2 and T3 for breast cancer T1 for cancer of the ovary (note: T2 tumours of the ovary are considered as ***advanced***)
2 =	A - Advanced disease *Advanced disease means:* T3 and T4 tumours for cancers of the stomach, colon, rectum, liver, lung, cervix, prostate T4 for breast (note: T3 tumours of the breast are considered as ***localised***) T2 and T3 for ovary
9 =	X – Cannot be assessed: no information on tumour size category

Variable 31 Condensed N **Optional**

Numeric variable, one digit.

Code	Meaning
1 =	N0 – No regional lymph node invasion by tumour
2 =	N+ – Regional lymph nodes invaded by tumour
9 =	NX – Cannot be assessed: no information on nodal status

Variable 32 Condensed M **Optional**

Numeric variable, one digit.

This code is based on the best available information - clinical, instrumental or pathological. Clinical signs and findings are sufficient to justify classifying a tumour as having metastasised (M+), even without *pathological* confirmation of metastatic deposits.

Code	Meaning
1 =	M0 – No distant metastasis
2 =	M+ – Distant metastases present
9 =	MX – Cannot be assessed: no information on whether metastases are present

Variable 33 Dukes' stage Optional

Numeric variable, one digit.

This variable is optional. If you choose not to supply it, please assign the code 9 to this variable in every tumour record.

Dukes' stage[22] is a specialised classification of tumour stage for cancers of the ***colon and rectum only***. For **all other index cancers**, please assign the code 9 to every tumour record.

The TNM classification is preferable, because it is more detailed.

Dukes' stage should ***only*** be reported if TNM stage data (variables 24-29) are not available.

Dukes' stage was later modified[23] to include a category for metastasis (group D), and sub-categories for direct extension in groups B and C. Modified Dukes' stage is no longer recommended for clinical use, but it is still widely used.

Code		Meaning
1	=	Dukes' stage A (this is equivalent to T1N0M0 or T2N0M0)
2	=	Dukes' stage B (this is equivalent to T3N0M0 or T4N0M0)
3	=	Dukes' stage C (this is equivalent to T(any)N1M0 or T(any)N2M0)
4	=	Dukes' stage D (this is equivalent to T(any)N(any)M1)
9	=	Dukes' stage missing: no information on Dukes' stage

Variable 34 FIGO stage Optional

Numeric variable, one digit.

This variable is optional. If you choose not to supply it, please assign the code 9 to this variable in every tumour record.

FIGO stage[24] is a specialised classification of tumour stage for **cervical, ovarian** and other gynaecological cancers. For **all other index cancers**, please assign the code 9 to this variable in every tumour record.

The TNM classification is preferable, because it is more detailed.

FIGO stage should ***only*** be reported if TNM stage data (variables 24-29) are not available.

FIGO stage provides five broad categories:

Stage 0	carcinoma *in situ* (common in cervical cancer)
Stage I	confined to the organ of origin
Stage II	invasion of surrounding organs or tissue
Stage III	spread to distant nodes or tissue within the pelvis
Stage IV	distant metastasis(es)

Cancer of the cervix (C530, C531, C538, C539)

Code Meaning

0 = FIGO Stage 0 – carcinoma *in situ*

1 = FIGO Stage I – Tumour confined to cervix (extension to corpus uteri should be disregarded)
This code should also be used for sub-categories IA, IA1, IA2, IB, IB1 and IB2

2 = FIGO Stage II – Tumour invades beyond uterus but not to pelvic wall or lower third of vagina
This code should also be used for sub-categories IIA, IIA1, IIA2 and IIB

3 = FIGO Stage III – Tumour extends to pelvic wall or lower third of vagina, or causes hydronephrosis
This code should also be used for sub-categories IIIA and IIIB

4 = FIGO Stage IVA – Tumour invades mucosa of the bladder or rectum, or extends beyond true pelvis

5 = FIGO Stage IVB – Distant metastasis

9 = FIGO Stage unknown

Cancer of the ovary (C569) or Fallopian tube (C570)

Code Meaning

1 = FIGO Stage I – Tumour limited to one or both ovaries*
This code should also be used for sub-categories IA, IB and IC

2 = FIGO Stage II – Tumour involves one or both ovaries* with pelvic extension
This code should also be used for sub-categories IIA, IIB and IIC

3 = FIGO Stage III – Tumour involves one or both ovaries* with microscopically confirmed peritoneal metastasis outside the pelvis, and/or regional lymph node metastasis
This code should also be used for sub-categories IIIA, IIIB and IIIC

4 = FIGO Stage IV – Distant metastasis outside the peritoneal cavity

9 = FIGO Stage unknown

*** For malignancies of the Fallopian tubes (C570), replace "ovaries" with "Fallopian tubes".**

Variable 35 Tumour size Optional

Numeric variable, from one to three digits.

Tumour size (maximum tumour diameter) must be reported in **millimetres**, as an integer.

For **breast cancer**, tumour size should be based on histological examination, if available.

For **lung cancer**, tumour size may be available by imaging.

It is difficult to be prescriptive about the maximum physical dimensions of a tumour. We will accept values in the range 1 - 300mm (1mm – 30cm). For example, a tumour with a maximum diameter of 35mm (3.5cm) would be coded as “35”.

Code		Meaning
1-300	=	maximum tumour diameter, **in millimetres**
999	=	maximum tumour diameter is not known, ***or***
		maximum tumour diameter is not applicable (**leukaemia**), ***or***
		this variable will not be supplied

Zero is not a valid tumour dimension. If no data are available for a solid tumour, please code this variable as 999.

Variable 36 Number of lymph nodes examined Optional

Numeric variable, one or two digits.

Report the exact ***number of lymph nodes examined***, as recorded in the pathological record: valid range 0-98.

This variable should be coded to 99 in ***all records*** if:
- you choose not to supply this variable, ***or***
- the data file is for **leukaemia**

For the nine **solid tumours**, this variable should be coded to 99 if:
- no information is available on the number of lymph nodes examined, ***or***
- no pathological examination was done

Variable 37 Number of lymph nodes involved Optional

Numeric variable, one or two digits.

Report the exact ***number of involved lymph nodes (lymph nodes containing tumour cells)***, as recorded in the pathological report: valid range 0-98.

This variable should be coded to 99 in ***all records*** if:
- you choose not to supply this variable, ***or***
- the data file is for **leukaemia**

For the nine **solid tumours**, this variable should be coded to 99 if:
- no information is available on the number of involved lymph nodes, ***or***
- lymph nodes *were* involved, but the *number* of involved lymph nodes is unknown, ***or***
- no pathological examination was done

4 Valid values for each variable

Name of variable	Short description	No. of digits or characters	Valid values (or *range* of valid values)	Value to be used when valid data are missing
VAR1	Country	4	See list on page 4	Not allowed
VAR2	Registry	1, 2 or 3	Use value provided*	Not allowed
VAR3	Person code	Up to 15	Numeric or alphanumeric	Not allowed
VAR4	Tumour code	Up to 10	Numeric or alphanumeric	Not allowed
VAR5	IARC CHECK flag[1]	1	1,2,3,4	9
VAR6	Sex	1	1,2	9
VAR7	Region[2]	Up to 5	Numeric (see page 10)	99999
VAR8	Race/ethnicity[2]	1 or 2	Numeric (see page 11)	99
VAR9	Day of birth	1 or 2	1-31	99
VAR10	Month of birth	1 or 2	1-12	99
VAR11	Year of birth	4	1895 - latest year	9999
VAR12	Day of diagnosis	1 or 2	1-31	99
VAR13	Month of diagnosis	1 or 2	1-12	99
VAR14	Year of diagnosis	4	1995 - latest year	9999
VAR15	Last known vital status	1	1,2,3	9
VAR16	Day of last known vital status	1 or 2	1-31	99
VAR17	Month of last known vital status	1 or 2	1-12	99
VAR18	Year of last known vital status	4	1995 - latest year	9999
VAR19	Basis of diagnosis	1	1,2,3,4,5,6,7,8	9
VAR20	ICD-O-3 Topography	4	Alphanumeric	Not allowed
VAR21	ICD-O-3 Morphology	4	8000-9989	9999
VAR22	Behaviour	1	0,1,2,3,6,9	Not allowed
VAR23	SEER Summary Stage 2000	1	0,1,2,3,4,5,7	9
VAR24	Pathological T	1	0,1,2,3,4,8	9
VAR25	Pathological N	1	0,1,2,3	9
VAR26	Pathological M	1	1	9
VAR27	Clinical T	1	0,1,2,3,4,8	9
VAR28	Clinical N	1	0,1,2,3	9
VAR29	Clinical M	1	0,1	9
VAR30	Condensed T	1	1,2	9
VAR31	Condensed N	1	1,2	9
VAR32	Condensed M	1	1,2	9
VAR33	Dukes' stage[3]	1	1,2,3,4	9
VAR34	FIGO stage[4]	1	0,1,2,3,4,5	9
VAR35	Tumour size	1, 2 or 3	1-300	999
VAR36	No. of lymph nodes examined	1 or 2	0-98	99
VAR37	No. of lymph nodes involved	1 or 2	0-98	99

***** The code to use for this variable will be supplied to you separately

[1] If you do ***not wish*** to supply this variable, please fill it with the number 9 in every record (see last column)

[2] If you ***do wish*** to supply this variable, you must agree the valid codes with us in advance; if you do ***not wish*** to supply it, please fill it with the number 99999 (VAR7) or 99 (VAR8) (see last column).

[3] Valid values ***only*** for colon and rectal cancer; for all other cancers, use the number 9 (see last column)

[4] Valid values ***only*** for cervical and ovarian cancer; for all other cancers, use the number 9 (see last column)

5 References

1. Fritz AG, Percy C, Jack A, Shanmugaratnam K, Sobin LH, Parkin DM, Whelan SL, eds. *International Classification of Diseases for Oncology (ICD-O)*. 3rd edn. Geneva: World Health Organisation; 2000
2. World Health Organisation. *International Classification of Diseases, 1975, 9th revision*. Geneva: WHO; 1977
3. World Health Organisation. *International statistical classification of diseases and related health problems. Tenth revision*. Geneva: WHO; 1994
4. Percy C, Van Holten V, Muir CS, eds. *International Classification of Diseases for Oncology (ICD-O), second edition*. Geneva: World Health Organisation; 1990
5. Woods LM, Rachet B, Ellis L, Coleman MP. Full dates (day, month, year) should be used in population-based cancer survival studies. *Int J Cancer* 2012; **131**: E1120-E1124
6. Sant M, Allemani C, Tereanu C, De Angelis R, Capocaccia R, Visser O, Marcos-Gragera R, Maynadié M, Simonetti A, Lutz J-M, Berrino F. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood* 2010; **116**: 3724-34
7. Marcos-Gragera R, Allemani C, Tereanu C, De Angelis R, Capocaccia R, Maynadié M, Luminari S, Ferretti S, Johannesen TB, Sankila R, Karjalainen-Lindsberg M-L, Simonetti A, Martos MC, Raphael M, Giraldo P, Sant M. Survival of European patients diagnosed with lymphoid neoplasms in 2000-2002: results of the HAEMACARE project. *Haematologica* 2011; **96**: 720-8
8. Maynadié M, De Angelis R, Marcos-Gragera R, Visser O, Allemani C, Tereanu C, Capocaccia R, Giacomin A, Lutz J-M, Martos C, Sankila R, Johannesen TB, Simonetti A, Sant M. Survival of European patients diagnosed with myeloid malignancies: a HAEMACARE study [epub ahead of print]. *Haematologica* 2012
9. Sant M, Karjalainen-Lindsberg M-L, Maynadié M, Raphael M, Ferretti S, Giacomin A, Tereanu C, Giraldo P, Marcos-Gragera R, Martos C, Lutz J-M, Visser O, HAEMACARE Working Group. Manual for coding and reporting haematological malignancies. *Tumori* 2010; **96 (Appendix)**: 1-32
10. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer* 2005; **103**: 1457-67
11. Sant M, Allemani C, Capocaccia R, Hakulinen T, Aareleid T, Coebergh JWW, Coleman MP, Grosclaude PC, Martinez-Garcia C, Bell CMJ, Williams EMI, Berrino F, EUROCARE Working Group. Stage at diagnosis is a key explanation of differences in breast cancer survival across Europe. *Int J Cancer* 2003; **106**: 416-22
12. Ciccolallo L, Capocaccia R, Coleman MP, Berrino F, Coebergh JWW, Damhuis RAM, Faivre J, Martinez-Garcia C, Møller H, Ponz de Leon M, Launoy G, Raverdy N, Williams EMI, Gatta G. Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut* 2005; **54**: 268-73
13. Sankaranarayanan R, Swaminathan R, Brenner H, Chen K, Chia KS, Chen JG, Law SCK, Ahn YO, Xiang YB, Yeole BB, Shin HR, Shanta V, Woo ZH, Martin N, Sumitsawan Y, Sriplung H, Barboza AO, Eser S, Nene BM, Suwanrungruang K, Jayalekshmi P, Dikshit R, Wabinga H, Esteban DB, Laudico A, Bhurgri Y, Bah E, Al-Hamdan N. Cancer survival in Africa, Asia, and Central America: a population-based study. *Lancet Oncol* 2010; **11**: 165-73
14. Maringe C, Walters S, Butler J, Coleman MP, Hacker N, Hanna L, Mosgaard BJ, Nordin A, Rosen B, Engholm G, Gjerstorff ML, Hatcher J, Johannesen TB, McGahan CE, Meechan D, Middleton R, Tracey E, Turner D, Richards MA, Rachet B, ICBP Module 1 Working Group. Stage at diagnosis and ovarian cancer survival: evidence from the International Cancer Benchmarking Partnership. *Gynec Oncol* 2012; **127**: 75-82
15. Allemani C, Sant M, Weir HK, Richardson LC, Baili P, Storm H, Siesling S, Torrella-Ramos A, Voogd AC, Aareleid T, Ardanaz E, Berrino F, Bielska-Lasota M, Bolick S, Cirilli C, Colonna M, Contiero P, Cress RD, Crocetti E, Fulton JP, Grosclaude P, Hakulinen T, Izarzugaza I, Malmström P, Peignaux K, Primic-Žakelj M, Rachtan J, Safaei Diba C, Sanchez M-J, Schymura MJ, Shen T, Traina A, Tryggvadóttir L, Tumino R, Velten M, Vercelli M, Wolf HJ, Woronoff A-S, Wu X, Coleman MP. Breast cancer survival in the US and Europe: a CONCORD high-resolution study. *Int J Cancer* 2013; **132**: 1170-81
16. Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: lessons from the International Cancer Benchmarking Partnership. *Int J Cancer* 2013; **132**: 676-85
17. Young JL, Roffers SD, Ries LAG, Fritz AG, Hurlbut AA. *SEER Summary Staging Manual - 2000: Codes and Coding Instructions. NIH Pub. No. 01-4969*. Bethesda, MD: National Cancer Institute; 2001
18. Sobin LH, Gospodarowicz M, Wittekind C. *TNM Classification of Malignant Tumours*. 7th edn. New York: John Wiley & Sons; 2009
19. Edge SB, Byrd DR, Carducci MA, Compton CC, Fritz AG, Greene F, Trotti A, eds. *American Joint Committee on Cancer - cancer staging manual*. 7th edn. New York, NY: Springer; 2009
20. Berrino F, Brown C, Moller T, Sobin LH, Faivre J. *ENCR recommendations: condensed TNM for coding the extent of disease*. Lyon, European Network of Cancer Registries, 2002.
21. Sobin LH, Wittekind C, eds. *TNM Classification of Malignant Tumours*. 6th edn. New York: John Wiley & Sons; 2002
22. Dukes CE. The classification of cancer of the rectum. *J Path Bacteriol* 1932; **35**: 323-32
23. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* 1954; **139**: 846-52
24. 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 Gyn Obs* 2000; **70**: 209-62