



ENCR Call for Data

21 May 2010

EUROCOURSE: data harmonisation, analysis and exchange.....	2
Annex 1. Submission instructions.....	3
Annex 2. Required data files	4
1. Cancer cases file	4
2. Population data file.....	8
3. Mortality data file	9
4. Population life tables	10
Annex 3. References.....	11
Annex 4. ENCR data call requirements at a glance.....	12

EUROCOURSE: data harmonisation, analysis and exchange

EUROCOURSE aims to strengthen the basis for monitoring cancer incidence and mortality in Europe through the provision of regular and timely information from European cancer registries, obtained via a single quality-assured gateway. The registries are equal partners in the project through their participation in the various working groups carrying out particular tasks or activities, as well as through the ENCR Steering Committee.

Work Packages 3 and 4 are concerned with improving the quality, coverage and use of cancer registration data in Europe (WP3) and the development, harmonisation, analysis and exchange of European cancer registry (WP4). Working Groups have been convened by the respective WP leaders, Roberto Zanetti (roberto.zanetti@cpo.it) and Harry Comber (h.comber@ncr.ie), to ensure, on behalf of the registries, appropriate strategies for the quality assurance, analysis and online dissemination of the datasets received. The recommendations of a Working Group chaired by Henrik Møller (henrik.moller@kcl.ac.uk) provided the basis for this call, and two further Working Groups will advise on 1) the analysis and presentation of key statistics on the cancer burden in Europe (chaired by Freddie Bray, Brayf@iarc.fr) and 2) the sharing of data and its ownership (chaired by Stefano Rosso, stefano.rosso@cpo.it).

It is envisaged that three levels of EUROCOURSE output will be considered by these working groups in close consultation with the Registries, the ENCR Steering Committee and its Secretariat at IARC:

EUCAN 2010

An online analysis tool for investigating the geographical patterns of cancer incidence and mortality according to the major sites of cancer. National estimates for 2010 will be derived for the 40 countries of greater-Europe.

EUROCAN

Using NORDCAN (<http://ancr.nu>) as a model, EUROCAN will provide tabular and graphical analyses of country-specific patterns and trends on the basis of national registry data or via an aggregation of regional data. The inclusion of sub-national data (from individual regional registries) will be explored by the above parties.

EUROCIM-2

A new version of EUROCIM comprising of databases of registry-specific incidence and mortality data available at different levels of topographical / morphological detail (ICD-10, histology-specific entities etc.). Initially delivered to all registries who submit their data to the gateway, issues of ownership, provision of analytic tools etc. will be explored by the above parties.

Annex 1. Submission instructions

Files preparation

The cancer cases file should be zipped and protected by a password before submitting. There are no specific requirements for the names of the files.

It is preferable NOT to zip together the other files (population, coding, etc.), so as to allow automating sorting of files and renaming them to a standard filenames on submission.

The following files should be submitted:

1. Cancer cases file
2. Population file
3. Life-table file (if follow-up information is available in the Cancer cases file)
4. Mortality file
5. A file specifying any non-standard coding variables

CIN portal at IARC

Submission of the files through the new CIN portal is strongly recommended. Using the portal will allow you benefit from the following features:

- High security and speed of data transfer
- No packaging
- No need for accompanying correspondence
- On-line confirmation of participation
- Automated data submission and its acknowledgement
- On-line monitoring of the progress of data processing at IARC
- One submission for several projects

Instructions are provided on the portal for each step required for the file upload. Specific help is available on request.

During the submission on the portal it will be possible to specify the call each data file is submitted for. If a registry wishes to submit different data items or different coding for different studies, a separate submission of relevant files is recommended for each study. The CIN portal is located at the following address: <http://cinportal.iarc.fr/>.

To enter the portal, the registry-specific username and password are required. They were communicated to each registry individually.

Contact and other submission paths

It is possible to submit the files on a CD or by e-mail to the address shown below, although the submission through the new portal is preferable:

ENCR Secretariat
150 cours Albert Thomas
69372 Lyon Cedex 08
FRANCE
encr@iarc.fr

Deadline for inclusion in the first database release: 30 June 2010

Annex 2. Required data files

The list below shows the requirements of the content and format of the files to be submitted, separately for each file type.

1. Cancer cases file

Inclusion criteria

Period

- All available registration years considered complete
- The latest available closing date of follow-up for vital status

Tumour

All primary tumours, including, if collected:

- Basal cell and squamous cell carcinomas of skin
- Non-malignant tumours of CNS
- Non-malignant tumours of the urinary bladder
- *in situ* tumours of breast, cervix, colon, rectum and melanoma of the skin

Age

- All ages eligible for registration
- In the age-restricted cancer registries, all subsequent primary tumours of the registered patients should be included irrespectively of age at diagnosis, where available

Requested quality and coding

- Data should be verified and corrected before the submission using, for example, IARCTools or DEPEdits.
- Duplicate registrations should be removed.
- All multiple primary tumours should be retained in the file.
- The coding schemes outlined in detail in Table 1, Example 1 and Example 2 should be used.
- In particular, ICD-O-3 coding system is requested for all relevant variables. IARCTools software could be used to convert coding of other systems.

Requested format

In general, the files should be formatted as follows:

- One record per line
- Delimited field format
- Predefined format and record layout outlined below
- Mandatory variables must be completed for all records
- Optional variables be completed for all records, if collected

Data for each registry (distinguished by its registry code) should be recorded in a separate file.

Requested accompanying information

Coding scheme for all variables using other than recommended coding (See Table 1)

Table 1. Cancer cases file: record layout and recommended coding
(F=Numeric, A=Alphanumeric)

Variable	Format	Recommended coding of values		Definition, notes
		Regular	Missing or Unknown	
1. Patient identification number*	A20		Not allowed	A unique identifier of the patient in the originating registry, which is not necessarily the official personal identification number.
2. Tumour sequence number*	F2	00=single tumour 01=1 st of more 02=2 nd of more 03=3 rd of more Etc.	99	A sequential number of all primary tumours diagnosed over the lifetime of a person in chronological order and included in the dataset.
3. Total number of tumours	F2	Number	99	A count of all tumours diagnosed in one registered patient during the registry's activity.
4. Date of birth*	F8	DDMMYYYY	99999999	If date of birth cannot be provided in full, the next possible/available detail should be included, such as MMYYYY or DDYYYY
5. Sex at birth*	F1	1=Male 2=Female	9	
6. Date of incidence*	F8	DDMMYYYY	9999YYYY	See ENCR recommendations: Date of incidence
7. Date of registration	F8	DDMMYYYY	99999999	Date when case first recorded in the registry database
8. Age in years*	F3		999	Latest completed year of age at the time of diagnosis
9. ICDO-3 Topography*	A4	ICD-O-3	Not allowed	Example: C531
10. ICDO-3 Morphology*	F4	ICD-O-3	Not allowed	Example: 8170
11. ICD-O-3 Behaviour*	F1	ICD-O-3	Not allowed	Example: 3
12. Basis of diagnosis*	F1	ICD-O-3		See also ENCR recommendations for coding basis of diagnosis
		0=DCO 1=Clinical 2=Microscopic	9	Minimum coding requirements
13. Detection of non-symptomatic cancer in an organised population-based screening programme	F1	1=Screen detected 2=Other	9	See ENCR recommendations: Method of detection in relation to screening. An 'organised screening programme' is defined as 'men and/or women in an identified population, invited to participate in a screening programme'.

* Mandatory variable

Variable	Format	Recommended coding of values		Definition, notes
		Regular	Missing or Unknown	
14. Incidental finding of cancer at the autopsy	F1	1=Yes 2=No	9	
15. Grade (ICDO-3)	F1	ICD-O-3	9	Arabic numerals should be used for all codes
16. Laterality of paired organ	F1	1=Unilateral 2=Bilateral 3=Right 4=Left	9	A bilateral tumour to be recorded in a single record
17. Vital status at last contact	F1	1=Alive 2=Dead	9	Coding of 'Lost to follow-up' should be as follows: 'Vital status' should be set to 1 (alive) and 'Date of end of follow-up' to the latest date when the patient was confirmed alive.
18. Date of end of follow-up	F8	DDMMYYYY	99999999	Equivalent to 'Date of death' for deceased patients and to 'Date of last contact' for those alive at the end of follow-up.
19. Age at end of follow-up	F3	Number	999	Integer part of the age reached at the date quoted in 18. Must be provided if any of 4, 6 or 18 is incomplete.
20. Duration of survival	F5	Number of days	99999	Days between '6. Date of incidence' and '18. Date of end of follow-up'. Must be provided if either 6 or 18 are incomplete.
21. Cause of Death	A4	ICD	9999	Code of the underlying cause of death
22. ICD edition	F2	06=ICD-6 07=ICD-7 08=ICD-8 09=ICD-9 10=ICD-10	99, Not allowed for records with 21 provided	Coding system used for the underlying cause of death
23. TNM stage, primary site	A4	AJCC/UICC	9999	Examples: TX, pT0, Tis, T1a, T2, pT3, T4
24. TNM stage, lymph nodes	A5	AJCC/UICC	99999	Examples: pNX, N0, pN1sn, N2b, N3
25. TNM stage, metastases	A4	AJCC/UICC	9999	Examples: M0, M1
26. TNM stage grouping	A2	AJCC/UICC	99	Arabic numerals should be used
27. TNM reference	A3	U06=UICC 6 th Edition U07=UICC 7 th Edition	999	
28. Condensed TNM, T	A2	T1=Localised T2=Advanced	TX	See ENCR recommendations: condensed TNM
29. Condensed TNM, N	A2	N1=No reg. nodes N2=Regional nodes	NX	See ENCR recommendations: condensed TNM
30. Condensed TNM, M	A2	M1=No dist. metast. M2=Distant metast.	MX	See ENCR recommendations: condensed TNM

Variable	Format	Recommended coding of values		Definition, notes
		Regular	Missing or Unknown	
31. Extent of disease	F1	1=Confined 2=Adjacent tissues 3=Distant organs 4=2 or 3 NOS 5=1 or 2 NOS	9	See EUROCORE 5 protocol
32. Tumour size in mm	F3	Millimetres	999	
33. Examined nodes	F2	Number	99	Number of nodes examined
34. Metastatic nodes	F2	Number	99	Number of nodes with the sign of malignant spread
35. C factor	F1	1=C1, standard dg. 2=C2, special dg.		See ENCR recommendations: condensed TNM
36. Surgery	F1	1=yes 2=no	9	Non-investigative surgery within 6 months of diagnosis
37. Systemic therapy	F1	1=yes 2=no	9	Systemic therapy (e.g. chemotherapy) within 6 months of diagnosis, including adjuvant systemic therapy
38. Radiotherapy	F1	1=yes 2=no	9	Radiotherapy within 6 months of diagnosis, including adjuvant radiotherapy
39. Hormone therapy	F1	1=yes 2=no	9	Hormone therapy within 6 months of diagnosis, including adjuvant hormonal therapy
40. Curative intent: surgery with curative intent	F1	1=yes 2=no	9	EUROCORE-5 protocol
41. Curative intent: chemotherapy, including adjuvant	F1	1=yes 2=no	9	EUROCORE-5 protocol
42. Curative intent: radiotherapy	F1	1=yes 2=no	9	EUROCORE-5 protocol
43. Curative intent: other therapy	F1	1=yes 2=no	9	EUROCORE-5 protocol
44. Symptomatic treatment	F1	1=yes 2=no	9	EUROCORE-5 protocol
45. IARC flag	F1	1=OK 2=OK after verification 0=failed	9	Indicator of successful pass through IARC checks and conversions

2. Population data file

Scope

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

- registration area
- time period
- age-range

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

File format

Ideally, the file should contain a population figure for every combination of:

- Calendar year
- Sex
- Age (year unit if possible, standard 18 age-groups otherwise)

Each line of the population file should include number of residents for the combination of calendar year, sex and age, as shown below (Example 1).

Alternatively, the next finest level of detail available should be provided (Example 2).

Data for each registry (distinguished by its registry code) should be recorded in a separate file or worksheet.

The codes for sex should be the same as in the cancer cases file.

Example 1:

Year	Sex	Age unit	Number of residents
1991	1	0	N _{1,1,0}
1991	1	1	N _{1,1,1}
1991	1	2	N _{1,1,2}
1991	1
1991	1	100	N _{1,1,100}
1991	2	0	N _{1,2,0}
1991	2	1	N _{1,2,1}
1991	2
1991	2	100	N _{1,2,100}
1992	1	0	N _{2,1,0}
1992	1
1992	1	100	N _{2,1,100}
1992	2	0	N _{2,2,0}
.....
2008	2	100	N _{18,2,100}

Example 2:

Year	Sex	Age group	Number of residents
1991	1	1	N _{1,1,1}
1991	1	2	N _{1,1,2}
1991	1	3	N _{1,1,3}
1991	1
1991	1	18	N _{1,1,18}
1991	2	1	N _{1,2,1}
1991	2	2	N _{1,2,2}
1991	2
1991	2	18	N _{1,2,18}
1992	1	1	N _{2,1,1}
1992	1
1992	1	18	N _{2,1,18}
1992	2	1	N _{2,2,1}
.....
2008	2	18	N _{18,2,18}

Accompanying information required

- Any other than recommended coding should be documented.
- Reference to the source of population data should be provided.

3. Mortality data file

It is not required to submit national mortality statistics, which may be retrieved directly from the WHO database.

Scope

The mortality data for the **sub-national regions** covered by a cancer registry should include all past 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
- age-range

File format

Ideally, the file should contain the number of deaths for every combination of:

- Calendar year
- Sex
- Age (year unit if possible, standard 18 age-groups otherwise)
- Cause of death (3 characters of ICD code)

Each line of the mortality file should include number of deaths for the combination of calendar year, sex, age and cause of death, as shown below (Example 1).

Alternatively, the next finest level of detail available should be provided (Example 2).

Data for each region (corresponding to the registration areas distinguished by separate registry codes) should be recorded in a separate file or worksheet.

The codes for sex should be the same as in the cancer cases file.

Example 3:

Year	Sex	Age unit	Cause of death	Deaths
1991	1	0	C00	D _{1,1,0,C00}
1991	1	1	C00	D _{1,1,1,C00}
1991	1	2	C00	D _{1,1,2,C00}
1991	1
1991	1	100	C00	D _{1,1,100,C00}
1991	2	0	C00	D _{1,2,0,C00}
1991	2	1	C00	D _{1,2,1,C00}
1991	2
1991	2	100	C00	D _{1,2,100,C00}
1992	1	0	C01	D _{2,1,0,C01}
1992	1
1992	1	100	C01	D _{2,1,100,C01}
1992	2	0	C01	D _{2,2,0,C01}
.....
2008	2	100	C97	D _{18,2,100,C97}

Example 4:

Year	Sex	Age group	Cause of death	Deaths
1991	1	1	140	D _{1,1,1,140}
1991	1	2	140	D _{1,1,2,140}
1991	1	3	140	D _{1,1,3,140}
1991	1
1991	1	18	140	D _{1,1,18,140}
1991	2	1	140	D _{1,2,1,140}
1991	2	2	140	D _{1,2,2,140}
1991	2
1991	2	18	140	D _{1,2,18,140}
1992	1	1	141	D _{2,1,1,141}
1992	1
1992	1	18	141	D _{2,1,18,141}
1992	2	1	141	D _{2,2,1,141}
.....
2008	2	18	208	D _{18,2,18, 208}

Accompanying information required

- Any other than recommended coding should be documented.
- Reference to the source of mortality data should be provided.

4. Population life tables

These files are only requested if the submitted cancer cases file contain the information on follow-up for vital status.

Format

Ideally, the file should contain probabilities for every combination of:

- Calendar year (all incidence and follow-up years)
- Sex
- Age (year unit)

Alternatively, the next finest level of detail available, with a maximum of 5-year age groups and a minimum of the 85+ last age group (or higher).

Survival probabilities for general (all causes) mortality should be provided to 6 decimal places or an equivalent number of significant figures.

Data for each registry (distinguished by its registry code) should be recorded in a separate file or worksheet.

Accompanying information required

- The layout and format of the file
- The key for any specific codes used for variables
- Method of smoothing if mortality probabilities are provided for 5 years periods (as opposed to single years)
- The source reference
- The reference person

Annex 3. References

Conversion and verification: rules and software

1. Ferlay J, Burkhard C, Whelan S, Parkin DM. CHECK AND CONVERSION PROGRAMS FOR CANCER REGISTRIES (IARC/IACR Tools for Cancer Registries). IARC Technical Report No. 42 Lyon, 2005
2. <http://www.iacr.com.fr/>, 'Software', 'IARCcrgTools'; accessed on 28 April 2010.
3. <http://www.enchr.com.fr/>, 'Downloads', 'DEPeditis'; accessed on 28 April 2010.

ENCR recommendations

4. Parkin DM, Tyczynski JE, Démaret E, eds. Standards and guidelines for cancer registration in Europe. IARC Technical Publication No 40, 2003: 69–73.
5. <http://www.enchr.com.fr/>, 'Downloads', 'Recommendations'; accessed on 28 April 2010.

EUROCARE 5

6. <http://www.eurocare.it/Protocols/>; accessed on 28 April 2010.

IACR standards:

7. <http://www.iacr.com.fr/>, 'Standards'; accessed on 28 April 2010.

ICD-10

8. WHO. International Statistical Classification of Diseases and Related Health Problems (10th Revision). Geneva: World Health Organisation, 1992.

ICD-O-3

9. Fritz A, Percy C, Jack A, et al, eds. International Classification of Diseases for Oncology (3rd edition). Geneva: World Health Organisation, 2000.

TNM classification

10. Sobin LH, Gospodarowicz MK, Wittekind C (Eds.) TNM Classification of Malignant Tumours, 7th Edition November 2009, Wiley-Blackwell.
11. www.uicc.org/tnm; accessed on 28 April 2010
12. <http://www.wileyanduicc.com/>; accessed on 28 April 2010.

Annex 4. ENCR data call requirements at a glance

Cancer cases file

- Verification of inclusion criteria
- Validation
- Conversion to ICD-O-3
- Removal of duplicate registrations
- Retaining all multiple primaries
- Checking mandatory and optional variables
- Adhering to the recommended coding
- Formatting
- Zipping and protecting

Other data files

- Population file (minimum submission)
- Mortality data base
- Population life tables

Other information

- Non-standard coding schemes used in any data files
- Reference to the source of population data
- Reference to the source of life tables

CIN portal at IARC

<http://cinportal.iarc.fr/>

Username

Password

Permissions for use of the data

Deadline: 30 June 2010