CHAPTER 2
Registration

Contents

CHAPTER 2
Registration

Reportable cases

Cases studied by a general population based cancer registry include all malignant tumors arisen in the observed population. Traditionally, according to the *9th Revision of the International Classification of Diseases*(ICD-9), diseases included between codes 140.0 and 208.9 have been considered reportable, with an extension in the 210.0-239.9 interval to include further cases of interest among benign cases, *in situ* cases, and cases of uncertain behavior (it is also possible to retrieve these through the ICD-9-CM procedure codes, selecting specific treatments). Using classifications with morphology fields (ICD-O, SNOMED) all codes of sections 8 and 9 with behavior 3 or higher have been considered, except intracranial and intraspinal tumors, where benign cases and cases of uncertain behavior (/0; /1) are also registered, and bladder tumors that also include forms of uncertain behavior and *in situ* (/1; /2) forms. Epidemiological reasons have gradually required registration of lesions at an earlier clinical-biological level, preceding invasion (according to the common meaning of ICD-O-3 a malignant behavior is the ascertained infiltration of a tumor with a consequent possibility of metastasis), as these are important in signaling emerging disease and, especially, for monitoring screening programs (see section on screening). This attention has especially focused on intraepithelial (*in situ*) forms of carcinomas and in some cases on dysplastic lesions that can be identified in the carcinogenic pathway of certain neoplasms (e.g., cervical cancer). Likewise, numerous exceptions to the rules of reportability of multiple primaries have become widespread, resulting in registration of synchronous or metachronous tumors in the same organ or in paired organs, especially in diseases that are subject to early diagnosis procedures or therapy evaluation. Further changes and add-ons to registration procedures have been made in relation to specific studies. Furthermore, the second and particularly the third edition of the ICD-O re-classified as malignant diseases previously classified as of uncertain behavior (e.g., all lympho-hematological diseases classified in the 270-289 interval of ICD-9).

Therefore, reportable cases include at least the following:

- all invasive tumors, usually included in the 140-208 interval of ICD-9, in the C00-C97 interval of ICD-10 and with M-8000–M-9989 morphology of ICD-O with /3 behavior;
- all intracranial and intraspinal tumors, regardless of their behavior;
- lympho-hematological diseases included in the 270-289 interval of ICD-9 and with behavior /3 of ICD-O;
- *In situ* tumors (ICD-9: 230-234; ICD-10: D00-D09; ICD-O-3: M-8000–M-9989 with behavior /2) of the bladder and other neoplasms subject to screening;
- tumors of uncertain behavior (/1) of the bladder and sites subject to screening;
- tumors of uncertain behavior (/1) which might be classified as NSE (*not sure eligibility*) and/or DCI.

Whereas the need to respect traditional reportability criteria (only malignant tumors) and registration rules (e.g., multiple primaries) for the production of incidence data, over time the difference in cases actually collected and those used for incidence calculation and other occurrence indicators has widened. This enables registries to fully respect international rules while at the same time using the same information flow to meet needs of further epidemiology research, in an increasingly positive trend in the relations between registries and reporting institutions.

Incident cases

By definition, incident cases are cases diagnosed for the first time within a particular time frame, subject to multiple primary validation rules.

The following are considered incident cases:

- Malignant/invasive neoplasms (ICD-9 140-208; ICD-10 C00-C97 and ICD-O morphology M-8000–M-9989 with behavior /3) with the freedom to choose not to include therein basal cell and squamous cell epitheliomas of the skin;
- DCO cases;
- *in situ* bladder tumors and bladder tumors of uncertain behavior;
- intracranial and intraspinal tumors, with benign or uncertain behavior where required by specific research protocols.[1]

More ample case sets are generally part of specific research projects and must in any case be explicitly described when presenting the data.

The evolving of registration rules over the years, particularly with respect to the definition of multiple primaries, has changed reportability criteria, resulting at times in obvious problems in analyzing historical cases

(e.g., skin tumors); likewise, the evolving of cancer classification systems has redefined categories and behaviors, so that cases that were previously not reportable are now reportable, and vice versa (e.g., myelodysplasias). This evolving in the comprehension of certain diseases and the way they are considered is an integral part of epidemiological research. The section devoted to specific cancer sites provides a detailed analysis of the main aspects of these changes and the possible registration problems they may give rise to. For instance, incidence assessment for breast carcinomas *in situ* appears necessary, considering their importance in the field of screening programs. For the same reason it is useful, when possible, to focus equal attention on pre-invasive cervical and colorectal lesions. Registration of these lesions must follow criteria of completeness (suggested in the handbook's section on specific anatomic sites), without obviously interfering with the indicators of incidence of invasive forms.

Residence

The characteristic which best defines the role and specificity of a population-based cancer registry is its capacity to refer incident cases to a well-identified denominator (at-risk population), represented by the population residing in a specific area.

A reference to registered residency is very precise thanks to the organization of municipal registers, which manage the demographic aspects of the population. Compared to other types of residency (temporary, usual but not official, etc.) registered residency identifies a more stable population, reducing potential bias when comparing different areas. The registered residence of a patient at the time of the first tumor diagnosis is therefore a condition for inclusion of the case in the incidence archive of an area; it must be verified at municipal registers or by reviewing archives (e.g., lists of Local Health Unit patients) which draw data directly from municipal registers.

Another problem is due to the time factor: there should be a close time correspondence between "at-risk resident population" and "incident resident cases"; actually, the resident data is measured with greater precision in the latter, the migration of which is also better known. Thus the temporal correspondence required between the two populations is a more or less strongly pursued quality goal rather than a fact. With respect to the "time factor," residency criteria required for cases are generally the following:

- *restrictive criterion:* being officially resident on the day of diagnosis;
- *broad criterion*: being officially resident in the year of diagnosis (at least: emigrated on the 1st of January or immigrated on the 31st of December).

Reasons behind the choice of one criterion over the other are more empirical than formal; however, the first is preferable and should be adopted, on the basis of the following considerations:

- the information systems on which linkages are based refer to registered residency;
- the current population covered in Italy is such that there exists a risk of duplicates.

A potential problem of the reference population is that the officially resident population and the population receiving care is not perfectly identical. The population receiving care generally is the population that actually resides in a given area, and it includes temporary residents, in certain cases augmented by migration flows tied to specific conditions of risk or particular care needs. Since knowledge of the epidemiological profile of these sectors of the population (included in the target population of screening programs) is very important, registration of new cancer diagnosis in these people is recommended; careful estimates of the number of patients who are not officially resident in a given area can provide the possibility of obtaining useful indicators for analysis and comparison with the resident population that is normally considered by the registries.

Laterality and paired organs

When no specific topography code is available for laterality (e.g., colon) a data field must be inserted to provide information about this aspect. Recording laterality is especially important in paired organs (lung, breast, kidney, etc.) as well as in a number of single organs that can be divided into lobes or sections (thyroid, prostate, breast, colon, pancreas), to allow for more detailed analysis of incidence of multiple recurrent tumors.

Registration of tumors that do not fall into the category of multiple primaries according to international rules (see below) can provide useful information for secondary prevention, especially for certain sites (e.g., breast). It must be borne in mind that, as explained in greater detail in the relevant section, according to international rules paired organs must be considered a single organ and morphologically homologous tumors must be considered only once for incidence calculations, independently of their subsequent onset in the contralateral organ.

Premalignant lesions

With the exception of bladder tumors, premalignant lesions used to be generally excluded from the primary objectives of a cancer registry. Their incidence, however, has gradually taken on considerable importance, in particular in certain sites (e.g., cervix, bowel, breast, prostate, skin melanoma) on which early diagnosis, both

spontaneous and organized in screening programs, focuses. Detailed description and suggestions on what information to report are provided in the section devoted to specific tumor sites.

Lesions of uncertain behavior

Lesions of uncertain behavior are not generally included in the analysis of incidence according to international rules (with the exception of bladder tumors and, in certain cases, intracranial and intraspinal tumors), but their trace in health files (HDDs, pathology reports, etc.) is useful for forms which may later reveal themselves as malignant (e.g., ovary), as well as an important additional source of information for diseases that subsequent classifications should consider to have malignant behavior (e.g., myelodysplasias in ICD-O-3).

Data sources

Registration in Italy is based on **active and passive casefinding**. Active casefinding is defined as a process in which all available sources are used to find new cases and report and classify them. Generally, registries identify an initial set of reportable cases by extracting data from a local information system (**passive casefinding**); additional sources such as hospital records, pathology reports, and death certificates are then used to complement the data (**active casefinding**). Current integration of various sources of information, including computerized sources, makes this subdivision difficult to identify nowadays; in effect, registries generally need to use active casefinding (visiting archives, hospitals, etc.) in the conclusive stage of definition of cases that are dubious or have insufficient information.

The acquisition of data from multiple sources obviously results in an increase in the quality of the information available, but on the other hand it increases the number of multiple reports for each single case, which must then be examined and rejected. This makes automatic procedures of record linkage of data referring to the same individual extremely useful for the reduction of the workload in this phase. On the other hand, different sources can present small differences in the personal identification variables that hinder matching procedures.

To correctly reconstruct prevalent cases, newly activated registries must needs have reporting sources that can provide data for at least two years prior to the beginning of registration.

Public register sources

Currently, health data management systems rely more and more on official public register sources (municipal registers), minimizing copying and data duplication, which are the cause of most errors in personal identification data. Obviously, one of the foremost tasks of registries is to ensure availability of an official public register (a town bureau or an agency directly linked to it), since it is essential to correctly identify patients throughout their clinical history.

The population data of municipal registers and offices of vital records should be privileged as the first-choice source. Except in the few cases when the data are automatically available and continuously updated even on the registry's data set, use of these sources is essential for later corrections, as well (during follow-up or during acquisition of ISTAT death records). It must be borne in mind that for subjects lost to follow-up, records of death are also entered at the town of birth.

Patient identification databases of the reference Local Health Units are to be considered incomplete, unless they are adequately linked with municipal registers or the Ministry of Finance (for the tax ID number) and regularly updated. They can be useful when health care services to a person cease (with motivations provided) or when a health tax exemption is issued (for NSE cases).

As a rule, personal identification data must not be changed, unless errors are found. The only particular case is that of town of residence and address, if registered (in the case of multiple primaries, they might vary more than once for the same patient). In any case, the town and address of residence registered at first incidence must not be changed, since they refer to the patient and not the case. Changes in residence can be registered, if so chosen with specific dates, in order to address follow-up to the correct reporting facility.

Essential personal data that registries must archive for each case are the following:

- surname;
- name;
- sex;
- date of birth;
- town of birth;
- town of residence;
- street address (optional data field);
- date of death;
- place of death;
- place of emigration (optional data field, but recommended for follow-up);
- date of emigration (optional data field, but recommended for follow-up);

Correct recording and archiving of these data is indispensable for all subsequent acquisition of information, in particular to provide the possibility of carrying out successful record linkage between different sources.

The most common problems concern data entry errors, both due to reporting sources and to the registries'

work, but other errors may occur due to different classification and interpretation systems. Below are a few examples of typical mismatches, along with suggestions of possible check procedures.

- *Complex surname* (presence of prepositions, stresses, apostrophes, double surname, foreign surname, married surname). In general, linkage works if the tax ID number is used (as an integration), thus reducing this type of error; however, many sources do not include the tax ID number.
- *Name*. A common type of error occurs when the case has a first name and a middle name; public register rules have changed in the past decade and they now exclude the middle name if a comma follows the first name, but patients tend to follow their habit when registering; in other cases it is the spelling as a single name or as two names that leads to mistakes (e.g., Gian Franco, Gian-Franco, Gianfranco). Other frequent types of error include misspelled foreign names, and names used both for males and females (e.g., Andrea) or with different stresses (Élia, Elìa) which can cause erroneous gender assignments.
- *Date of birth*. Errors in the data entry stage are possible, especially when the data are copied (typos); other more common problems depend on the date format (Italian, American) used by different systems (when outputs of a program are used as cascade inputs for the next phase) and, frequently, by the different re-formatting of the date (century digits) which make passages from one application to another (Excel, Access, etc.) at high risk for errors (generating missing cases).
- *Age*. The issues which may obviously arise due to date problems are compounded, in automatic computation of age at diagnosis, by rounding up and down of ages, which can lead to misadjustments between age classes calculated in different situations (usually truncation to age in completed years is preferable).
- *Place of birth*. Place of birth, along with surname, name, sex and date of birth, are the essential minimum data set for personal identification. Some problems might be caused by the automatic coding systems of the town (e.g., HDDs). The following issues may be encountered:
  - towns that became part of a newly established or another province; in some cases, codes remain as they originally were;
  - towns that definitively or temporarily ceased to exist as such (having been joined with others and then split again, or having changed name); they are often not recognized by the system, sometimes they are assigned to the "foreign" category, or in any case given wrong attribution (even the list provided by the Ministry of Health does not include a great part of these variations, which are however reported by the Ministry of Finance in the official program for building the tax ID number; in some cases ISTAT assignments are obviously wrong); as a further complication, a new revision procedure for tax ID numbers began last year, with assignment of the code do discontinued town names;
  - countries that changed name or jurisdiction (see previous situation).
- *Town of residence*. Besides the cases of towns that were included into a new province or another province, maintaining their original codes in some data sets, the following types of error are also common:
  - the provincial capital is assigned as the town of residence;
  - classification is wrong due to incomplete name of the town (e.g., Villanova D'Asti assigned to Villanova Biellese);
  - the place were the person lives but is not officially resident is assigned as the town of residence, or the name of a suburb is assigned instead of the name of the town (generally in hardcopy documents).
- *Tax ID number*. Since it is automatically reconstructed, it contains errors on the date of birth and the town of birth; however, the surname-name pair alone is too generic to enable linkage; the problem of "homocodes" (the same tax ID number is assigned to two citizens and is subject to change by the Ministry of Finance when discovered) appears to have minor relevance.

The AIRTUM website and its version on CD-ROM provides a complete listing of all towns and countries with indication of its original and updated ISTAT code and its tax ID digits, with a reconstruction of the history of the codes where possible (registries may propose the addition of further information); In some cases, it was not possible to update the ISTAT code to the new situation because data were not available.

With the aid of the tax ID number, the place of birth can be reconstructed, and the differences between tax ID numbers found in different dataset can be explained: in the past few years, many health registers reconstructed tax ID numbers based on discontinued town names (thus a subject born in 1920 in Baggio, which became part of Milan in 1923, may present code A545 for Baggio or code F205 for Milan).

Mandatory sources of information

Adequate levels of completeness and precision in registration can be achieved by registries, including for accreditation purposes, if they make use at least of the following sources:

- hospital discharge data (HDD);
- clinical records;
- archives of anatomical pathology, histology, and cytology results;
- mortality archives.

Additional sources of information

Based on local situations and needs, besides the above-mentioned essential sources, every registry generally has one or more complementary sources:

- childhood cancer registries (if present in the same area covered by the registry);
- hospital-based and organ-based registries (if present in the same area covered by the registry);
- regional boards for authorization of cost reimbursement for care received abroad;
- regional mesothelioma registries;
- national disease registries (e.g., bone cancer, retinoblastoma);
- palliative care units, hospices, home care provided as part of a public service;
- freestanding radiation centers;
- diagnostic medical imaging centers;
- independent laboratories;
- independent oncology units;
- screening centers;
- health tax exemption archives for patients suffering from cancer;
- public health care general practice doctors and pediatricians;
- network of pediatric oncologists.

Critical aspects of current sources of information

Hospital discharge data (HDD)

The hospital discharge data chart summarizes the basic information regarding patients' hospitalization. It is divided into three sections: one identifies patients and their position in the national healthcare system, one gives information on the hospitalization process (ward and date of admission, transfers, type and ward of discharge), and the last includes data on diagnosed and treated diseases. This data flow, which has been electronically available regionally since the 1990s, is a primary reporting source which is generally easy to access and enables registries to reconstruct historical data for each patient, including extra-regional hospitalizations. The HDD archive is a primary source for computerized registries, and it is essential for manual and semi-automated registries to identify the clinical records that need to be reviewed.

The following critical issues must be underlined: there may be errors in the personal data entered, or encrypting errors made by some regions; disease and treatment coding may be incomplete or erroneous; changes in hospital identification codes may have been implemented; the clinical record may have been closed before the diagnostic process was over (in the case of surgical procedures, the histological diagnosis might be available after the clinical record has been closed, and therefore the record will not contain cancer codes). Therefore any casefinding procedure or record linkage on HDDs must aim for maximum sensitivity, even if this reduces specificity, and must include additional disease codes and a separate search on surgical procedures; furthermore, it can be useful to set up different association procedures in the record linkage (for instance: surname-name; surnamename-date of birth; surname-date of birth; different strings of digits from the tax ID number).

The Appendix presents a software program for the selection of disease and procedure codes.

Clinical records

Clinical records are an essential source of information, both in the casefinding stage and during data review, especially with respect to the exact definition of the incidence date and to the availability of remote pathology reports (with obvious reference to multiple primaries, prior surgical procedures, etc.); if the patient's record includes a complete collection of information on the disease (date and place of diagnosis and treatment, site and histological type, this diagnostic information must be used to register the case and the incidence date.

Whereas on one hand the quality of clinical records has improved, on the other hand many diagnostic procedures which previously took place in a hospital setting are now performed on an outpatient basis.

For hospitalizations found through HDD flows, that are "external" to the area covered by the registry, systematic casefinding should be carried out, particularly at major Italian cancer centers, to avoid selective loss of important cases; registries will establish how often casefinding audits will be done, based among other things on the need to optimize case ascertainment costs.

Archives of anatomical pathology, histology, and cytology results

These archives, too, are among the primary essential sources of information for every registry, as they provide irreplaceable information with respect to both nosology

and disease extension (staging), and the cancer's date of incidence. When archives are computerized, record linkage operations are possible. Pathology reports have become an essential source with the increase of diagnostic and therapeutic activity and follow-ups, representing furthermore the only resource that can be used to retrieve autopsy data confirming a diagnostic suspicion that was not resolved in lifetime, or to identify cancer with no clinical evidence.
The following problems may be encountered: personal identification that is incomplete and inadequate for record linkage; absence of staging in the report; use of terms that can lead to erroneous classification, for instance in the distinction between benign, borderline, and malignant cases that require an initial agreement on diagnostic definitions; unavailability of explicit diagnosis (e.g., "non-Hodgkin's lymphoma" is listed without the type, or data on lymph nodes are missing); the description of a single lesion is spread over more than one report.

Mortality archives

Since 1996, Local Health Units (ASL) have maintained registries of causes of death, archiving the data contained in ISTAT death records, which can sometimes be organized by greater areas (large city, province, region).
Availability of these data is essential for all registries (and cannot be replaced by data provided by ISTAT, which as a rule do not contain names and surnames) since, integrated with incidence data, they enable registries in particular:

- to complete the follow-up on the life status of registered patients (at municipal registers), gathering data on demises (date, place, cause of death);
- to compute correct prevalence measures and estimates;
- to find cases previously overlooked by incidence recording and trace them back (if necessary using any indications on the disease's duration), estimating these losses using DCI (death certificate initiated) and DCO (death certificate only) indicators
- to solve NSE cases (see Chapter 3);
- correct any mistakes in vital records.

Childhood cancer registries

When childhood cancer registries are present in the area covered by the general registry, casefinding should preferably be carried out through the childhood cancer registry, to avoid the burden of double registration for the main pediatric oncology facilities.

Hospital-based and disease-based registries

These archives are useful to integrate data flow, especially with respect to the periods preceding registry start-up, to check prevalent cases and for research purposes.

Regional committees for authorization of cost reimbursement for care received abroad

Their data must be systematically reviewed especially in the case of diseases for which many patients seek care at internationally renowned centers.

Mesothelioma registries

National disease registries (e.g., bone cancer, retinoblastoma)

Palliative care units, hospices, home care provided as part of a public service

Where they exist, these facilities gather case sets of patients who often are not treated or do not undergo invasive biopsy procedures due to the advanced stage of the disease. These case sets enable registries to solve part of NSE cases; they should also be considered for trace-back in DCI cases. Obviously, they take on particular importance for studies on the type of care oncology patients receive.

Freestanding radiation centers

It is useful to review these archives for the periods preceding registry start-up, to verify prevalent cases. In several instances they can provide missing data on staging and therapy.

Diagnostic medical imaging centers

Reviewing diagnostic imaging reports can prove useful to solve uncertain cases (NSE) and in trace-backs, in particular for diseases with low histological confirmation (lung, liver, pancreas, intracranial neoplasms).

Independent laboratories

Laboratory reports can be useful to solve uncertain cases (NSE) and in trace-backs, in particular for hematological diseases and for diseases in which specific markers play an important diagnostic role.

Independent oncology units

Data that are available at oncology services are particularly important to reconstruct the history of patients treated surgically at other facilities and with clinical records that are difficult to retrieve. Information on pre-treatment clinical staging (if it complements pathological staging) is essential for all studies with clinical impact.

Screening centers

Thanks to the exchange of information with screening centers, registries can assess the completeness of their data and enrich the amount of specific information for each case; these centers, on the other hand, can

complement the registry's archive with their patients' screening history and carry out an evaluation of the program's coverage. A specific section of this handbook deals with the relationship between registries and screening centers.

Health tax exemption archives for patients suffering from cancer

These archives are useful to find cases that had escaped notice, or as a support in identifying the incidence date. They are a specific source, but possibly one with low sensitivity (for patients who have other type of health coverage or exemptions).

Network of pediatric oncologists

The network of pediatric oncologists, which cooperates with childhood cancer registries, can be useful for the definition of cases treated in centers that cannot be directly accessed.

Public health care general practice doctors and pediatricians

Cooperation with general practitioners and pediatricians is especially useful in "second level" checks for cases that do not receive sufficient documentation from primary sources.

Finally, the performance of a clinical test or a treatment on a patient on a certain date obviously implies the fact that the patient was alive at the time; availability of these sources can be useful also for follow-up completion, in cases where it is difficult to acquire information from municipal registers.

Record linkage systems

Due to the ever increasing availability of computerized archives of the main data sources (HDDs, pathological reports, etc.) even non automated registries must needs activate record linkage systems to allow casefinding and subsequent unification of the various pieces of information referring to a single case. Linkage systems can vary based on archive complexity, their number, available computer resources, and required performance.
The best subject identification procedures have several hierarchical linkage keys, since, for instance, tax ID numbers, which are often reconstructed from the data, are affected by any errors in the data used to generate them; linkage keys are built by grouping personal data or part of them together (e.g., last name, initial of first name, date of birth, etc.) taking care to remove spaces, apostrophes, and stresses from both first and last name. Unless specific needs require otherwise, casefinding must aim for maximum sensitivity, even if this reduces specificity (for instance, in the case of HDDs this means including additional disease codes or performing separate searches on surgical procedures). Analysis of missing links should also be used to improve linkage keys and casefinding criteria.
Appendix 2 presents a software program produced by the Aviano CRO. The software, which has already been used in the CARL study, matches cases based on vital records using a probabilistic approach, which therefore makes it possible not to miss cases in which writing errors have created distortions in the personal data.

Data collection

Standard layout

An AIRTUM technical workgroup has created a record layout that is the basic structure of the national database.[2] The set of required data fields includes the variables that are internationally defined as essential[3,4] and provides a common reference for all Italian registries, making all information provided by registries immediately available and usable for national studies.
For some diseases, such as breast cancer, shared research projects that include additional clinical-biological data fields, which also follow international recommendations, have already been implemented.

Incidence

Malignant tumors incident in the period considered are collected, on the base of the classification systems in use at the time the case is defined. Tumors registered and defined as benign, uncertain, and *in situ* are also gathered, on the base of the classification used by the registry at the time the case is defined (this applies to registries that routinely collect non malignant tumors, as well). For the purpose of defining multiple primaries, the 2004 IARC/IACR/ENCR classification is used.[5]
The entire available incidence period must be submitted. Each new submission must include, besides the new incidence data, all cases incident in the preceding years (with any corrections and changes added since the previous submission), regardless of the fact that they have already been submitted.

Incidence record layout

In text formats, the number of digits shown in Table 1 (page II-16) represents the length of the data field in the layout. Data fields must be entered in sequence and their position in every record must be constant. In database formats the layout shown in Table 1 must be followed with respect to the type of data field. Further variables may be collected and implemented in the dataset that is submitted to the database. In particular, we recommend use of specific data fields to identify the following:

- case codes previously used, when a variation in coding system has occurred;

- patient codes previously used, when a variation in the coding system has occurred;
- a specific classification code for the case ("TIPO CASO" data field), using the following terminology:
  - C = true incident case
  - M = "missing" compared to the data provided to IARC
  - D = DCO
  - N = patient not resident at diagnosis, case arising during the period of observation
  - X = repeated cases, according to IARCcrg Tools - multiple routine
  - R = NSE or DCI cases still undefined (temporary value, not to be included in the data set to be submitted to the Database);
- a specific code (data field "CASO MULTIPLO") which makes it possible to distinguish malignant cancer events subsequent to the first tumor in patients with multiple primaries; this is because to measure survival only the first malignant tumor is considered (or in the case of the bladder the first tumor with any behavior) and in any case not a DCO case and not autopsy detected, unless the first tumor is an epithelial tumor of the skin; the code must be used when the data set does not provide another way to recognize these cases;
- the TNM stage, according to the rules described in the paragraph on TNM staging.

A copy of every data set submitted to the database or to IARC must naturally be archived and kept at the registry without further modifications, to enable later checks or ascertainments.

Mortality

The **individual data** referring to deaths **from all causes** occurring among residents in the area covered by the registry are collected. Individual data is needed to enable use of the SEER*Stat software (see Chapter 5). When it is not possible to report individual data, we recommend to submit aggregated data using data fields and formats that have been previously agreed on with the database managers. Preference is accorded to ISTAT mortality; if it is not available, local mortality can be used in areas where it is available and reliable. Causes of death can be acquired through linkage with the ISTAT national archive.

Mortality must be provided for the registry's entire area of coverage and for the entire period available. Each new submission must include, besides the new mortality data, all deaths that have occurred in the preceding years, regardless of the fact they have already been submitted.

Mortality record layout

In text formats, the number of digits shown in Table 2 (page II-18) represents the length of the data field in the layout. Data fields must be entered in sequence and their position in every record must be constant. In database formats the layout shown in Table 2 must be followed with respect to the type of data field.

Populations

Population data referring to the years for which the registry has contributed to the database with incidence and mortality data are gathered. Populations for the entire period of available incidence and mortality must be submitted. Each new submission must include, besides the new population data, all populations for the preceding years, regardless of the fact they have already been submitted.

Population record layout

In text formats, the number of digits shown in Table 3 (page II-19) represents the length of the data field in the layout. Data fields must be entered in sequence and their position in every record must be constant. In database formats the layout shown in Table 3 must be followed with respect to the type of data field.

A closer look at certain essential data fields

Case identifier

Two codes must be added to the previously listed data fields, to enable the case to be identified unequivocally and anonymously, so that data can be shared in multicentric studies while safeguarding the privacy of patients. The first of these codes identifies the patient, and it must be used without variation for each report of incidence, making it possible to identify (multiple) tumors of the same subject and all consequent correlated data (prevalence, survival, etc.); the second code must obviously identify unambiguously the single tumor. The combination of the two codes must be able to yield correct identification and assignment of each lesion at any time.

Incidence date

The definition of the date of incidence is a crucial moment in the review of documentation available for each case, since distortions may be introduced when defining the initial onset of a neoplasm, with repercussions on its staging, on patient survival, and on the comparability of the criteria followed by each registry.

The date of the hospital admission that led to the diagnosis of cancer was for a long time the most commonly used date suggested as date of incidence. Due to the current organization of health services,

especially following the large-scale diffusion of minimally invasive and highly efficient diagnostic procedures (e.g., CT, MRI, endoscopy, and ultrasound- and radio-guided aspiration biopsy), cancer diagnosis ever more frequently precedes hospital admission, while hospitalization is used to provide therapy and disease staging.

All these considerations, which have been debated at length internationally,[6] led to the choice of one of the following diagnostic events in the clinical history of patients as incidence date:

- date of pathological examination (qualitatively the most accurate and reliable);
- date of clinical diagnosis (cancer diagnosis must be definitive and not merely expressed as a doubt);
- date of hospital admission during which the first diagnosis of tumor is made;
- other dates (therapies, death, etc.).

According to international recommendations, the criteria according to which the date of diagnosis is chosen can be deduced considering the following classification in order of (decreasing) importance: the first available date is given preference. If an event with higher priority is recorded **within three months** of the date initially chosen, it must be considered as the date with higher priority.

The following must be taken in consideration, in decreasing order of importance:

a. date of the first histological confirmation or "certain" cytological confirmation of the neoplasm (with the exception of autopsy microscopic confirmation); the date must refer to the arrival of the specimen at the pathology laboratory;
b. date of hospital admission during which the first diagnosis of tumor is made;
c. when the diagnosis is not made during hospitalization: date of the first clinical or laboratory examination in which the diagnosis of tumor is made;
d. dates other than those given in a, b, c;
e. date of death, should the only information available be death of the patient by cancer (DCO), or when the cancer diagnosis is only made during autopsy, without prior clinical findings (in this situation the case is excluded from survival analysis).

Whatever the available date of incidence, it obviously can never be posterior to the beginning of a cancer-specific treatment, or the decision not to treat the patient, or the patient's death.

The choice of incidence date (date of the first type of diagnosis) does not necessarily decide the data field "basis of diagnosis," which must in any case provide the highest diagnostic level in terms of certainty, reached during the patient's entire observed history.

Metachronous tumors must be registered with their own incidence date, following the rules for multiple primaries (see Chapter 3). Metachronous tumors preceding the start-up of a registry, if found, must be included in the archive, but must not be considered for incidence; for survival analysis, subsequent tumors are not considered, regardless of their reportability as multiple primaries, unless the first tumor is not malignant or it is an epithelial skin tumor.

Basis of diagnosis

Table 4 (page II-20) describes the rules for basis of diagnosis classification for the AIRTUM Database; we present it here as useful reference. To improve sensitivity on diagnostic sources, even if necessary only for the sites of greater interest, another classification can be used, which can differentiate the type of diagnostic examination performed and, in the case of histology or cytology, the type of procedure (histology from needle biopsy, surgical biopsy, endoscopic biopsy, or autopsy; exfoliative cytology, fine-needle aspiration cytology; hematology, blood or bone marrow). Finally, it must be remembered that cytogenetic and molecular diagnoses, to be considered higher in quality than microscopic histology, fall into codes 6, 7, and 8. Problems and particular issues in the assignment of this variable are discussed in greater detail in Chapter 3.

Topography and morphology coding

As established internationally, the ICD-O-3 coding must be considered official and mandatory both for topography and morphology, since it allows for the greatest possible adherence to the current disease classifications used in clinical and histological, cytological, and pathological diagnostics. Therefore, as a rule, neoplasms should be coded directly in ICD-O-3. For more detailed information on ICD-O-3 classification rules, please refer to the official ICD-O-3 text.[7] The impact of the rules on registration is discussed in the following chapters.

Extension and staging

Registration of extension and staging of tumors at diagnosis follows internationally published recommendations (ENCR). Extension and staging are indispensable data for clinical studies and assessment of screening impact. The recorded stage must come from official documents (anatomical pathology reports, clinical records, oncology reports). Any exception to this rule (reconstructions made by the registry on the basis of other documentation) must be mentioned when

presenting the data. In any case, staging must strictly follow the internationally agreed rules, which can be found in the following paragraph.

The TNM classification

The extension of neoplasms must be reported following the rules of the TNM system.[8] The TNM classification cannot be used for the staging of lymphomas, leukemia, or brain cancer; staging for these types of neoplasms is discussed in the section on these diseases. Code M = 1 is used to indicate distant metastasis evidence.

Staging is based on the best level of diagnosis in the initial stage of the disease and it is used to choose the best therapy plan. TNM staging can be based on postsurgical histopathological examination (pTNM); in the absence of surgical treatment, it is based on the clinical examinations which led to diagnosis (cTNM).

Other prefixes provide the context in which staging was performed.

- ”y” is used for cases in which classification is performed during or following initial multimodality therapy that might be capable of modifying the course of the disease; depending on the kind of classification ypTNM or ycTNM are used. A typical case is that of surgery preceded by radiation therapy or radio-chemotherapy (rectum, oral cavity, etc.) The finding of metastasis during the first cycle of treatment does not influence the initial staging.
- ”a” indicates that classification is first determined at autopsy (aTNM).
- ”r” identifies recurrent tumors, staged after a disease-free interval (rTNM). This must not be confused with staging of residual tumor after treatment; this type of staging is defined, similarly to the T classification, as RX-R0-R1 (microscopic residual tumor)-R2 (macroscopic residual tumor).
- ”m” is used to indicate the presence of multiple primary tumors at a single site (mTNM).

It is important to also consider the following aspects:

- registries must provide, if at all possible, the best level available, therefore pTNM;
- in most cases pathological staging refers only to T and N, while assessment of metastasis is clinical; in these cases it is preferable to distinguish between pTNM and pTNcM;
- to ensure comparability of data from different centers, cases in which staging was made in the clinical stage and is present in the original documentation must be distinguishable from cases where staging was made by the registry; if staging was performed by the registry it must be explicitly reported: an additional data field may be used for this purpose;
- in any case it is preferable for registries to stage a tumor as Tx and/or Nx and/or Mx, rather than forcing staging in the absence of sufficient evidence (as often occurs with elderly patients in whom invasive examinations are not a reasonable option), or on the basis of elements obtained at a later date than diagnosis or even following treatments, such as staging during follow-up; the presence of exhaustive stagings out of the total number of cases should be assessed as a critical element.

“Condensed” TNM

Should documentation on complete TNM staging not be available, and in the case that research protocols allow for it, registries may assign a level of disease extension at diagnosis according to the following “condensed” TNM classification, proposed by ENCR in 2002:[9]

T =	L (localized)	A (advanced)	X (not available)
N =	0	+	X
M =	0	+	X

where “T” and “N” are based, if possible, on the pathological report, or, subordinately, on clinical diagnosis (endoscopy, ultrasonography, radiography, etc.). The “M” factor can be based on the best information available (clinical, instrumental,, or pathological) or even on typical signs and symptoms.

The values of T corresponding to the “localized” and “advanced” categories and a detailed list of “advanced” T categories are shown in Tables 5 and 6 (pages II-21, II-22).

Stage N+ refers to regional lymph node involvement. Table 7 (page II-23) lists the definition of regional lymph nodes for every anatomic site.

As Table 7 shows, to correctly define T and N for some primary sites, a detailed subsite is required.

If the primary site is unknown (code C80.9 in ICD-O-3) T and N cannot be assigned; whereas stage M+ can, in any case, be automatically assigned. The “X” value for condensed TNM should be reserved for cases in which no information can be gleaned from any element in the entire documentation. If “X” is assigned as a result of reviewing the pathology report, an “L” or “A” code can be assigned on the basis of clinical documentation only (if available). For N or M staging, the “X” value must be assigned when there is no reasonable evidence of negativity. For instance, code N0/M0 (instead of NX/MX) can be assigned when no positive regional lymph nodes are found in a surgical specimen, or in the case of endoscopic resection of an intestinal tumor.

Tumors classified as non-resectable must be classified with M+ cases even in the absence of evident metastasis, because their expected outcome is analogous to that of metastatic neoplasms. This classification makes it possible to distinguish non-resectable cases from resected cases in which access to the histopathology report is not available.
The final result of "condensed" TNM staging is therefore as follows:

- localized (TL/N0/M0)
- local spread (TA/N0/M0)
- regional spread (every T/N+/M0)
- advanced stage
  - metastatic (every T/every N/M+)
  - non-resectable (Mx)(*except prostate cancer)*
- unknown extension (TX/NX/MX)

For breast, cervical, and colorectal cancer, registries must seek to obtain the most detailed staging possible (complete pTNM), which is the most suited to meeting assessment needs of screening programs.

Further information on extension and staging

Size of the tumor

It is often a decisive data item for the definition of stage T, but in many neoplasms it allows for more precise extension detail, which is just as important to assess diagnostic sensitivity (e.g., breast cancer with active screening programs) Every registry can decide for which cancers to add this parameter, and introduce an additional data field, usually using measurement in millimeters which should be found in the pathology report (maximum size of tumor; in skin melanoma the thickness is measured) or, as a second-choice source, in diagnostic procedure reports (x-ray, ultrasound scan, etc.).
If measurements have been taken on both a fresh and a fixed sample and they differ, the size of the fixed specimen must be recorded. In the case of multifocal and multiple synchronous tumors in one organ, the tumor with the highest category should be classified; the multiplicity or the number of tumors should be indicated in parentheses, e.g., T2(m) or T2(5). In synchronous bilateral cancers of paired organs, each tumor should be classified independently. In tumors of the thyroid, liver, ovary, and Fallopian tube, multiplicity is a criterion of T classification.

Number of sampled and metastatic lymph nodes

It is an additional parameter that significantly improves the sensitivity of N staging. In the case of N0 tumors, the number of total lymph nodes examined takes on an important role in staging quality control. An additional variable should be used to indicate N staging performed using sentinel lymph node assessment.

Certainty of data

The C-factor, or certainty factor, is offered by the TNM handbook as a further element of assessment of the validity of the information used for classification (Table 8, page II-26). Clinical TNM classification is therefore equivalent to certainty grade C1, C2, or C3, while pathological TNM classification is equivalent to level C4.
If condensed TNM is used, code C can be simplified as follows:

C1: evidence from standard diagnostic means (inspection, palpation, standard radiography, intraluminal endoscopy);
C2: evidence obtained by special diagnostic means: imaging (radiographic imaging in special projections, tomography, ultrasonography, lymphography, angiography, scintigraphy, MRI), endoscopic biopsy, cytology;
Cp: evidence based on post-surgical histopathological report

Further information

Specific studies may require further information complementing staging, or additional measurements in particular situations:

- positive margins;
- vascular invasion;
- lymphatic embolization;
- perineural invasion;
- quantification of tumor mass in biopsies (e.g., tumor length in cores and number of positive cores in prostate biopsies);
- presence of residual tumors;
- post-therapy staging (symbol: yTNM);
- staging of recurrence after disease-free interval (symbol: rTNM);
- autopsy staging (symbol: aTNM).

Disease staging

A further method of staging is Disease Staging (D.S.) which uses a system of classification that also measures the clinical severity of cases, producing clusters of patients that have similar prognosis and require similar treatment. For instance, in gastrointestinal tumors, where surgical treatment may be urgent (due to occlusions, perforations, hemorrhage, etc.), clinical severity of the case may justify significant differences in short-term survival of patients with the same stage (for further information see: www.medstat.com; www.diseasestaging.it; http://www.hcup-us.ahrq.gov/db/nation/nis/DiseaseStagingV5.2ReferenceGuide.pdf).

Treatment

The indication of the type of therapy generally provides further information in studies on the clinical course of the

disease and patient survival, usually in relation to prognostic, biological, and clinical factors. Due to the great number of possible therapeutic approaches for cancer it is not usually possible to register a limited number of standard variables as for other case features; generally studies that include analysis of types of treatment design study-specific detailed schemes, working closely with clinicians (surgeons, radiotherapists, oncologists, etc.)
However, a concise indication of the type of therapy adopted can be useful when analyzing large sets of numbers referring to treatment in screening programs or as indicators of changes related to new trends (e.g., conservative surgery) tied to early diagnosis. In these cases it can be useful to document the type of initial therapy undergone by the patient (usually undertaken no longer than four months after diagnosis), providing an indication of whether or not one of the main types of therapy was performed and whether it was completed. For instance, variables might include surgery, type of surgery, radiotherapy, chemotherapy, and hormone therapy.
If information on treatment is required, it is in any case necessary to record the most reliable and exhaustive data source to complete any information already available or run quality control checks on the data obtained.

Interaction with special registries

Special registries generally fall into one of the following two categories:

- patient-driven (childhood cancer registries, at-risk population registries, etc.);
- disease-driven (organ and disease registries, screening registries, etc.).

Cooperation between general and special registries is first of all important for sharing and exchanging data. Combining the two different types of information flow has the potential to ensure the best coverage: the general registry can offer its systematic reporting sources of current data (area coverage), while the special registry can provide important integration and verification on single cases (diagnostic detail, follow-up protocols, migration pathways, clarification on problematic or undefined cases, etc.).
Besides this type of cooperation, aimed at exploiting and maximizing the effectiveness of the two data flows, special and general registries can cooperate productively in various fields. Special registries can provide information that is more difficult to find for general registries, whereas the latter can offer broader possibilities of scientific research on issues which the rapid evolution and spreading of new diagnostic procedures and treatments have made urgent to address for health policy management. The study of risk factors and of their diffusion, biological characterization of neoplasms and its impact on prognosis and treatment, study of the appropriateness and accessibility of care all greatly profit from this type of interaction.

Screening programs

The traditional classification of "first identification methods" (incidental presentation, clinical symptoms) has lost over time much of its usefulness and applicability due to the growing importance of pre-clinical diagnosis, thanks both to screening tests (taken freely by individual patients or as part of organized screening programs), and to the increase in the sensitivity of laboratory and clinical diagnostic procedures. One of the functions of cancer registries is to assess organized screening programs. Screening programs for the prevention of breast cancer, cervical cancer, and colorectal cancer are now widespread and consolidated in most of Italy, although a great difference between screening coverage in the Center-North of the country and the South still exists.
Registries must therefore ensure that they are able to do the following:

- document the diagnostic method of incident cases, distinguishing between cases diagnosed after positive screening test; the proportion of screen-detected cases in the screening program's target age group is an index of the program's coverage;
- more broadly, document the screening history of every incident case and classify it appropriately, in order to also identify "interval" cases (cases diagnosed in the interval between two screening tests);
- work with screening programs to assess their impact on the target population;
- provide the distribution by stage of incident cases.

Breast, cervical, and colorectal screening is effective if the following long-term results are observed:

- a reduction in breast cancer mortality;
- a reduction in cervical cancer incidence and mortality;
- a reduction in colorectal cancer incidence and mortality;.

Achieve all this, however, requires time; therefore, for timely assessment of program effectiveness, early indicators should be considered, such as increase in survival, increase in early detection, and, in particular, the decrease of tumors in an advanced stage, bearing in mind that an increase in incidence is likely in the first screening rounds, due to subclinical cases discovered earlier thanks to the screening program (lead time).
Registries must therefore:

- be able to identify screen-detected cases in the database: this information can be retrieved through record linkage with the archive of the screening

program (this is the best method since this information is not always reported in routinely reviewed documents);
- retrieve information derived from the surgical pathology report, integrating and completing any prior diagnostic biopsy report;
- integrate routinely collected data with additional data for screened tumors;
- record *in situ* cancers of the breast;
- record, where possible, high grade dysplasia and *in situ* cancer of the cervix) (see Chapter 4).
- record, where possible, premalignant lesions (adenomas and high grade dysplasia) of the bowel;
- record, without however including them in incidence computation, contralateral *in situ* and invasive cancers of the breast; the reason for this is that screening programs focus on single lesions, while international cancer registration rules consider the breast a paired organ.

It can be useful to devote a specific chart for every tumor subject to screening, including variables with an impact on prognosis. The following is additional information that registries must record for correct assessment of stage distribution at diagnosis and for interpretation of survival differences:

- breast cancer:
  - patient's screening status (e.g., screen-detected, not screen-detected in woman with previous negative test, not invited, not responding to invitation);
  - TNM stage;
  - size (in mm) of lesion;
  - whether there is an *in situ component (with percentage);*
  - grade of invasive and *in situ component;*
  - type of surgery on the breast (biopsy, nodulectomy, wide excision, quadrantectomy, mastectomy);
  - whether sentinel lymph node and/or axillary dissection was performed;
  - number of lymph nodes sampled and number of metastatic lymph nodes, if any;
  - any cytological/histological examinations prior to surgery and their result (needle aspiration biopsy and needle biopsy);
  - focus of lesion (unifocal, multifocal, multicentric);
  - surgical margin status;
  - presence of vascular invasion;
  - neoadjuvant therapy, if performed;
  - biological characterization of lesions, if available (receptors, proliferation, oncogens);
- cervical tumors:
  - patient's screening status;
  - TNM stage;
  - FIGO stage;
  - number of lymph nodes sampled and number of metastatic lymph nodes, if any;
  - previous viral infections;
  - CIN/invasive tumor;
  - therapy performed;
- bowel tumors:
  - patient's screening status;
  - TNM stage;
  - Dukes stage (AJCC);
  - grade of differentiation;
  - number of lymph nodes examined and of metastatic lymph nodes, if any;
  - presence of carcinomatosis;
  - type of surgery;
  - surgical margin status.

(More detailed information on tumors subject to screening can be found in Chapter 4, "Specific tumor sites").

It is important to bear in mind that the production of these data must be timely: data usefulness is poor when more than 2-3 years go by.

Registries operating in areas covered by screening programs must cooperate closely with the screening program both locally and regionally; after defining together the information of interest (starting from the suggestions offered in this handbook), layouts and coding tables, specific protocols must be designed to regularly exchange information. The specific data thus collected will enable registries to build indicators, as well as to measure and analyze interval cancers, which according to European guidelines represent a measure of screening quality. On the other hand, screen-detected case reports can be useful to measure registries' completeness. For the indication of patient screening status, we suggest use of the classification listed in Table 9 (used in Emilia-Romagna, page II-27).

When this level of detail is not available, registries may preferably refer to the following classification (Italian multicentric study IMPATTO):

1) screen-detected cases;
2) not screen-detected in women with previous negative screening test (interval case);
3) cases arisen in women who did not answer screening invitation;
4) cases arisen in women who were never invited;
5) unknown or not applicable.

Or, as a less preferable option, registries may refer to the ENCR 2001 Recommendation,[11] which identifies at least four categories:

1) screen-detected cases;
2) interval cases (according to the program's definition, giving the period from the last test);

3) other cases;
4) unknown or not applicable.

Registry work also provides a means to measure the diffusion and impact of spontaneous screening, as proven, for instance, by the recent increase in incidence of prostate cancer, for instance, due to the widespread use of prostate specific antigen (PSA) testing.

References

1. European network of cancer registries. *Recommendations for Brain and Central Nervous System Tumours*. Lyon, 1998.
2. Paci E, Crocetti E, Buzzoni C et al. *Protocollo operativo per la gestione della Banca dati AIRTUM*. AIRTUM, 2005.
3. Jensen OM, Parkin DM, MacLennan R et al. *Cancer Registration, Principles and Methods*. IARC Scientific publications no. 95, Lyon, 1991.
4. European network of cancer registries. *Recommendations for a Standard Dataset for the European Network of Cancer Registries*. Lyon, 2005.
5. International Agency for Research on Cancer, World Health Organization, International Association of Cancer Registries, European Network of Cancer Registries. *International Rules for Multiple Primary Cancers (ICD-O third edition)*. Internal report no. 2004/02. Lyon, IARC 2004 (www.iacr.com.fr/ Mprules_july2004.pdf).
6. European network of cancer registries. *Working Group on Data Definition, Report: Incidence Date*. Lyon, 1995.
7. Fritz A, Percy C, Jack A et al. *International Classification of Diseases for Oncology*. Third edition. Geneva: World Health Organization, 2000.
8. Sobin LH, Witteking C. *TNM Classification of Malignant Tumors*. Sixth edition. New York, Wiley Liss, 2002.
9. European network of cancer registries. *Recommendations on Condensed TNM for the Extent of Disease*. Lyon, 2002.
10. Tyczynski JE, Démaret E, Parkin DM. *Standards and Guidelines for Cancer Registration in Europe*. IARC Technical publications no. 40, Lyon 2003.
11. European network of cancer registries. *Recommendation on Methods of Detection in Relation to Screening*. Lyon, 2001.

Tables

Table 1. AIRTUM Database incidence record layout BACK

Field name	Field length	Field type	Description
IDRT	2	text	registry ID: assigned centrally
IDPZ	34	text	patient ID: data field length, based on the longest entry
IDCASO	7	text	tumor ID: this field, coupled with the patient ID field, is the matching data pair which must remain unique and unvaried in time (multiple primaries)
SESSO	1	text	sex: valid values: 1 = male, 2 = female, 3 = other or unknown. Value 3 is not accepted by the check and must be used only after further verification.
GGNASC	2	number	day of birth
MMNASC	2	number	month of birth
AAAANASC	4	number	year of birth (four-digit format)
DATANASC	10	date	date of birth: (preferably) instead of the above fields, in the dd/mm/yyyy format
COMNASC	6	text	town of birth: ISTAT code (use the reference file in the Appendix). *This field has 6 digits: the shorter codes must be aligned right filling any spaces to the left with 0s which is why this must be considered a text field*
COMRES	2	text	town of residence: ISTAT code (use the reference file in the Appendix)
GGINCI	2	number	day of incidence
MMINCI	2	number	month of incidence
AAAAINCI	4	number	year of incidence (four-digit format):
DATAINCI	10	date	incidence date: (preferably) instead of the above fields, in the dd/mm/yyyy format
GGINSE	2	number	day case was entered: completeness control procedure
MMINSE	2	number	month case was entered: completeness control procedure
AAAAINSE	4	number	year case was entered: completeness control procedure
DATAINSE	10	date	date case was entered: completeness control procedure; (preferably) instead of the above fields, in the dd/mm/yyyy format
ICDO1T	5	xxx.x text	ICD-O-1 topography
ICDO1M	6	xxxx.x text	ICD-O-1 morphology+behavior
GICDO1	1	text	ICD-O-1 grading
ICD9	5	xxx.x text	ICD-9
ICDO2T	5	xxx.x text	ICD-O-2 topography
ICDO2M	6	xxxx.x text	ICD-O 2 morphology+behavior
GICDO2	1	text	ICD-O-2 grading
ICDO2COD	1	text	ICD-O-2 code: specify whether the ICD-O-2 codes were entered by the registry (0) or produced by a transcoding program (1)
ICD10	5	xxx.x text	ICD-10
ICD10COD	1	text	ICD-10 coding: specify whether the ICD-O-10 codes were entered by the registry (0) or produced by a transcoding program (1)
ICDO3T	5	xxx.x text	ICD-O-3 topography: must be provided in any case (whether original or transcoded)
ICDO3M	6	xxxx.x text	ICD-O-3 morphology+behavior: must be provided in any case (whether original or transcoded)
GICDO3	1	text	ICD-O-3 grading
ICDO3COD	1	text	ICD-O-3 code: specify whether the ICD-O-3 codes were entered by the registry (0) or produced by a transcoding program (1)
ICCC	6	xxxx.x text	*International Classification of Childhood Cancer*: to be used for cancer in subjects younger than 15 years of age; it can be transcoded using *Child-CHECK* ICD-O-1 and ICD-O-2
ICCCCOD	1	text	ICCC code: specify whether the ICCC codes were entered by the registry (0) or produced by a transcoding program (1)

BASE	1	text	basis of diagnosis: 0 = death certificate only (DCO) 1 = clinical only 2 = clinical investigations only 4 = specific tumor markers (see section on tumor markers) 5 = positive cytology 6 = positive histology on metastasis 7 = positive histology on primary 9 = unknown Note: to enable distinction between post-mortem examinations, code: 3 = autopsy without histology (otherwise coded as "2") 8 = autopsy with histology (otherwise coded as "6" or "7")
STATO	1	text	life status: 1 = alive; 2 = deceased 3 = lost to follow up
GGFOLLO	2	number	day of *follow-up*
MMFOLLO	2	number	month of *follow-up*
AAAAFOLLO	4	number	year of *follow-up*
DATAFOLLO	10	date	date of *follow-up*: (preferably) instead of the above fields, in the dd/mm/yyyy format
CAUSA9	5	xxx.x text	ICD-9 cause of death
CAUSA10	5	xxx.x text	ICD-10 cause of death (if available)
DCI	1	text	*Death certificate initiated*: case not initiated by death certificate (0) or case initiated by death certificate (1)
CHECK	1	text	IARC check (Chapter 5): before being submitted, data in ICD-O-3 must be run through the IARC *check* (IARCcrgTools downloadable from http://www.iacr.com.fr; DEPedits downloadable from http://www.ENCR.com.fr/download.htm); the programs at the moment have no updated version of the *Child -check*. Codes: 1 = check performed, 2 = check not performed (e.g., non malignant tumors)
VERIFI	1	text	verification for specific studies: for many years now many registries have cooperated in the EUROCARE project. To participate in this project, registries must run additional checks in addition to the standard IARC checks. The site/morphology combinations that need to be checked and corrected or confirmed are the following: ❖ basal cell carcinoma arising from the lip (ICD-9 140) ❖ basal cell adenocarcinoma (ICD-O-3 M-8147.3), accepted only with salivary glands (ICD-9 142); other sites requiring verification are ❖ villous adenocarcinoma (ICD-O-3 M-8262.3), accepted only with stomach (ICD-9 151), small intestine (ICD-9 152), colon (ICD-9 153), rectum (ICD-9 154), and gallbladder (ICD-9 156); other sites require code checks: 1 = not subject to EUROCARE checks; 2 = record checked, data confirmed.

Table 2. AIRTUM Database mortality record layout BACK

Field name	Field length	Field type	Description
IDRT	2	text	registry ID: this variable is assigned centrally
SESSO	1	text	sex: accepted values: 1 = male; 2 = female
AAAAMOR	4	number	year of death (four-digit format):
ETA	3	number	age in completed years
COMNAS	6	text	town of birth: ISTAT code. This field has 6 digits: the shorter codes must be aligned right filling any spaces to the left with 0s which is why this must be considered a text field
COMRES	6	text	town of residence: according to ISTAT classification
CAUSA9	5	xxx.x text	ICD-9 cause of death
CAUSA10	5	xxx.x text	ICD-10 cause of death: if available

Table 3. AIRTUM Database population record layout BACK

Field name	Field length	Field type	Description
IDRT	2	text	registry ID: this variable is assigned centrally
SESSO	1	text	sex: accepted values 1 = male; 2 = female
YEAR	4	number	year of residence (four-digit format):
ETA	3	number	year of age: preferable
CLETA	2	text	age in classes: only if year of age is not available. Classes must have five-year range, if possible separating age “0” (0, 1-4; 5-9, 10-14, etc.) up to class “85+”; if the last class is different from "85+" this must be reported when submitting the data
COMRES	6	text	town of residence: according to ISTAT classification (use the reference file in the appendix)
NUMBER	8	number	number of subjects: the population data must be submitted in aggregated form and the number of residents must be provided for each group

Table 4. Basis of diagnosis BACK

Code	Description	Criteria
0	DCO (death certificate only)	the available information comes from a death certificate
no microscopic diagnosis		
1	clinical	diagnosis was made prior to death, but without any of the following codes (codes 2-7)
2	clinical and laboratory examinations	all diagnostic methods (including radiodiagnostics, endoscopy, diagnostic imaging, ultrasound), with no subsequent histopathological examination, exploratory surgery
3	autopsy without histology	macroscopic diagnosis (this category, not included in the IARC/IACR table, was added to separate the data of direct observation of the tumor during autopsy)
4	specific tumor markers	diagnosis was made using biochemical and/or immunological markers that are specific for a certain cancer site
microscopic diagnosis		
5	cytology	analysis of cells taken from a primary or secondary cancer site (including needle biopsy and analysis of fluid collected by endoscopic drainage); it also includes microscopic examination of peripheral blood and bone marrow aspirates
6	histology of metastasis	histological examination performed on metastasis, including autopsy specimens
7	histology of primary	histological examination on primary tumor tissue, retrieved in any way, including any excision technique and bone marrow biopsy; also includes autopsy specimens of the primary
8	autopsy with concurrent or previous histology	this category, not included in the IARC/IACR table, was added to separate the data of direct observation of the tumor during autopsy
9	unknown method of diagnosis	

Table 5. Conventional values of T corresponding to $T_{localized}$ and $T_{advanced}$ BACK

Site	Localized	Advanced
lip and oral cavity	T_1-T_2	T_3-T_4
pharynx	T_1-T_2	T_3-T_4
larynx	T_1 -T_2	T_3-T_4
paranasal sinuses	T_1 -T_2	T_3-T_4
salivary glands	T_1-T_2	T_3 -T_4
thyroid	T_1-T_3	T_4
esophagus	T_1-T_2	T_3-T_4
stomach	T_1-T_2	T_3-T_4
small intestine	T_1-T_2	T_3 -T_4
colon and rectum	T_1-T_2	T_3 -T_4
anal canal	T_1-T_2	T_3-T_4
liver	T_1-T_2	T_3-T_4
gallbladder	T_1-T_2	T_3-T_4
extrahepatic bile ducts and hepato-duodenal ampulla	T_1-T_2	T_3
pancreas	T_1-T_2	T_3 -T_4
lung	T_1-T_2	T_3-T_4
pleura	T_1-T_2	T_3 -T_4
bone	T_1	T_2
soft tissues	T_1	T_2
skin	T_1-T_3	T_4
melanoma	T_1-T_3	T_4
breast	T_1-T_3	T_4
vulva	T_1-T_2	T_3-T_4
vagina	T_1-T_2	T_3-T_4
cervix	T_1-T_2	T_3 -T_4
uterus	T_1-T_2	T_3-T_4
ovary	T_1	T_2-T_3
Fallopian tube	T_1	T_2-T_3
trophoblast	T_1	T_2
penis	T_1-T_2	T_3-T_4
prostate	T_1-T_2	T_3-T_4
testicle	T_1-T_2	T_3-T_4
kidney	T_1-T_2	T_3 -T_4
renal pelvis and ureter	T_1-T_2	T_3 -T_4
bladder	T_1-T_2	T_3-T_4
urethra	T_1-T_2	T_3-T_4
eye	T_1-T_3	T_4
exception: sarcoma of the orbit	*T_1-T_2*	*T_3-T_4*

Table 6. Conventional definitions of T corresponding to $T_{advanced}$ BACK

Site	Minimum criteria $T_{advanced}$	Description
lip and oral cavity	T3	❖ tumor with greatest diameter >4 cm
pharynx, including base of tongue, soft palate, and uvula		
oropharynx	T3	❖ tumor with greatest diameter >4 cm
nasopharynx	T3	❖ tumor invading the bone or paranasal sinuses
hypopharynx	T3	❖ tumor with greatest diameter >4 cm or fixation of hemilarynx
larynx		
supraglottic	T3	❖ tumor confined to larynx with vocal chord fixation and/or invasion of the following: postcricoid area/pre-epiglottic tissues/paraglottic space/thyroid cartilage
glottic	T3	❖ tumor limited to the larynx with vocal chord fixation, involvement of the paraglottic space and thyroid cartilage
subglottic	T3	❖ tumor limited to larynx with vocal chord fixation
paranasal sinuses		
maxillary	T3	❖ see TNM handbook
ethmoidal	T3	❖ see TNM handbook
salivary glands (parotid, submandibular, sublingual)	T3	❖ tumor with greatest diameter >4 cm or extracapsular extension
thyroid	T4	❖ extracapsular extension or anaplastic histological type
esophagus	T3	❖ tumor extending beyond the tunica muscularis
stomach	T3	❖ tumor breaching the serosa (visceral peritoneum)
small intestine		
	T3	❖ tumor extending beyond the tunica muscularis
colon and rectum	T3	❖ tumor extending beyond the tunica muscularis
anal canal	T3	❖ tumor with greatest diameter >5 cm
liver (and intrahepatic bile ducts)	T3	❖ multiple lesions with diameter >5 cm or involvement of a main branch of the portal vein or the hepatic veins
gallbladder	T3	❖ tumor invades the serosa (visceral peritoneum) or adjacent structures
extrahepatic bile ducts	T3	❖ tumor invades adjacent structures: liver, pancreas, duodenum, gallbladder, colon, stomach
hepato-duodenal ampulla	T3	❖ tumor invades the pancreas or adjacent structures (the duodenal wall is considered T2)
pancreas	T3	❖ tumor is not confined to the pancreas
lung, mesothelioma	T3	❖ see TNM handbook
pleura	T3	❖ see TNM handbook
bone	T2	❖ tumor with greatest diameter >8 cm
soft tissues	T2	❖ tumor with greatest diameter >5 cm
skin carcinoma (except eyelid, vulva, and penis)	T4	❖ tumor invades deep extradermal structures (cartilage, skeletal muscle, bone)
malignant skin melanoma (except eyelid)	pT4	❖ tumor with thickness >4 mm
breast	T4	❖ any size with invasion of the thoracic wall or skin
vulva	T3	❖ tumor extends beyond vulva or perineum (urethra, vagina, anus/rectum, bladder)
vagina	T3	❖ tumor extend to pelvic wall and beyond
cervix	T3	❖ tumor extends beyond the uterus to the pelvic wall or the lower third of the vagina or beyond, hydronephrosis or kidney not functioning
uterus	T3	❖ tumor invades the serosa or extends beyond the uterus
ovary, Fallopian tube	T2	❖ tumor extends to pelvis
gestational trophoblastic disease	T2	❖ tumor extends beyond the uterus
penis	T3	❖ tumor invades the urethra or prostate
prostate	T3	❖ tumor has extracapsular extension
testicle	pT3	❖ tumor invades the spermatic cord
kidney	T3	❖ tumor extends beyond the kidney
renal pelvis, ureter	T3	❖ tumor extends beyond the tunica muscularis
bladder	T3	❖ tumor invades the perivesical tissue
urethra	T3	❖ tumor extends beyond the corpus spongiosum, prostate, or periurethral muscle
eye	T4	❖ see TNM handbook
sarcoma of the orbit	T3	❖ see TNM handbook

Table 7. Definition of "regional" lymph nodes BACK

Site	Regional lymph node stations
lip and oral cavity	cervical
pharynx (including base of tongue, soft palate, and uvula)	cervical
larynx	cervical
paranasal sinuses	cervical
salivary glands (parotid, submandibular, sublingual)	cervical
thyroid	cervical and sub-/supramediastinal
esophagus	
cervical	❖ scalene, anterior deep cervical (internal jugular), superior and inferior cervical, periesophageal, supraclavicular
intrathoracic	❖ superior periesophageal (above the azygos vein), subcarinal, inferior periesophageal (below the azygos vein), mediastinal, perigastric (excluding the celiac)
stomach	❖ perigastric along the lesser and greater curvature, along the left gastric artery, the common hepatic artery, the splenic artery and the celiac arteries, hepatoduodenal lymph nodes
gastroesophageal junction	❖ pericardial, left gastric, celiac, superior phrenic, inferior mediastinal paraesophageal
small intestine	
duodenum	❖ pancreaticoduodenal, pyloric, hepatic (pericholedochal, cystic, hilar), superior mesenteric
jejunum-ileum	❖ mesenteric (including superior mesenteric), ileocolic (only for the terminal ileum, including posterior cecal)
bowel	
appendix	❖ ileocolic
cecum	❖ ileocolic, right colic
ascending colon	❖ ileocolic, right colic, middle colic
hepatic flexure	❖ right colic, middle colic
transverse colon	❖ right colic, middle colic, left colic, inferior mesenteric
splenic flexure	❖ middle colic, left colic, inferior mesenteric
descending colon	❖ left colic, inferior mesenteric
sigmoid colon	❖ sigmoidal, left colic, superior rectal (hemorrhoidal), inferior mesenteric, sigmoid mesenteric
rectum	❖ superior, middle, and inferior rectal (hemorrhoidal), inferior mesenteric, internal iliac, mesorectal (pararectal), lateral sacral, presacral, sacral promontory (Gerota's)
anal canal	perirectal, internal iliac, inguinal
liver (including intrahepatic bile ducts)	hilar, hepatic (along the common hepatic artery), periportal (along the portal vein), along the inferior vena cava over the renal veins (except inferior phrenic lymph nodes)
gallbladder, extrahepatic bile ducts	pericholedochal, cystic duct, hilar, peripancreatic (head only), periduodenal, periportal, celiac, superior mesenteric ducts
hepato-duodenal ampulla	
superior	❖ superior to head and body of the pancreas
inferior	❖ inferior to head and body of the pancreas
anterior	❖ anterior pancreaticoduodenal, pyloric, and proximal mesenteric
posterior	❖ posterior pancreaticoduodenal, of the common hepatic duct, and proximal mesenteric
pancreas	
superior	❖ superior to head and body of the pancreas
inferior	❖ inferior to head and body of the pancreas
anterior	❖ anterior pancreaticoduodenal, pyloric (head only), proximal mesenteric
posterior	❖ posterior pancreaticoduodenal, of the common hepatic duct, proximal mesenteric
splenic	❖ splenic hilum and tail of the pancreas (body and tail only)
celiac	❖ for tumors of the head
lung, pleura mesothelioma	supradiaphragmatic stations: intrathoracic, scalene, of the aorta, internal mammary (only for pleural mesothelioma), supraclavicular

bone	of the primary tumor site (lympho node involvement is rare); if the lymph node status is not considered, they must be classified as N0 and not as Nx
soft tissues	of the primary tumor site (lympho node involvement is rare); if the lymph node status is not considered, they must be classified as N0 and not as Nx
skin (unilateral, including melanomas)	
head, neck	❖ preauricolar, submandibular, cervical and unilateral supraclavicular
thorax	❖ unilateral axillary
arm/shoulder	❖ epitrochlear, unilateral axillary
abdomen, loins, buttocks	❖ ipsilateral inguinal
leg/hip	❖ popliteal and ipsilateral inguinal
anal margin and perianal skin	❖ ipsilateral inguinal
- boundary zones (4 cm-wide bands are considered boundary zones)	regional lymph nodes are bilateral boundary zone between: zone along: right/left midline head and neck/thorax clavicula-acromion-upper shoulder blade edge thorax/arm shoulder-axilla-shoulder thorax/abdomen, loins, buttocks *front:* middle between navel and costal arch *back:* lower border of thoracic vertebrae (midtransverse-axis) abdomen, loins and buttock/leg groin-trochanter-gluteal sulcus
breast	1. axillary ipsilateral: interpectoral (Rotter nodes), along the axillary vein and its tributaries; they can be divided into the following levels: I level (lower axilla): lymph nodes lateral to the lateral border of pectoralis minor muscle II level (middle axilla): between the medial and lateral margins of the pectoralis minor muscle, interpectoral (Rotter nodes) III level (apex of axilla): medial to the medial margin of the pectoralis minor muscle, subclavicular, infraclavicular, apical 2. ipsilateral infraclavicular (subclavicular) 3. internal mammary (ipsilateral): lymph nodes in the intercostal spaces along the edge of the sternum in the endothoracic fascia 4. supraclavicular (ipsilateral)
Vulva	femoral, inguinal
vagina	
superior 2/3	❖ pelvic, including obturator, internal iliac (hypogastric), external iliac, pelvis NOS
inferior 1/3	❖ inguinal and femoral
cervix	paracervical, parametrial, hypogastric (internal iliac, obturator), common and external iliac, presacral, lateral sacral
body of uterus	pelvic (hypogastric – obturator, internal iliac – common and external iliac, parametrial and sacral), paraortic
ovary, Fallopian tube	hypogastric (obturator), common and external iliac, lateral sacral, paraortic, inguinal
gestational trophoblastic disease	the N classification is not taken into account
penis	deep and superficial inguinal, pelvic
prostate	lesser pelvis (below where the common iliac arteries fork) *(Note:: laterality does not affect N classification)*
testicle	paraortic abdominal (periaortic), preaortic, interaorto-cavali, precaval, paracaval, retrocaval, retroaortic, along the internal spermatic vein, intrapelvic and inguinal (only in the case of scrotal or inguinal surgery) *(Note: laterality does not affect N classification)*
kidney	hilar, abdominal paraortic and paracaval *(Note: laterality does not affect N classification)*

renal pelvis, ureter	hilar, abdominal paraortic and paracaval, intrapelvic (ureter only) *(Note: laterality does not affect N classification)*
bladder	lesser pelvis (below where the common iliac arteries fork) *(Note:: laterality does not affect N classification)*
urethra	inguinal, pelvic *(Note: laterality does not affect N classification)*
opthalmic cancer	preauricolar, submandibular, cervical
brain	TNM not applicable
Hodgkin's lymphomas	TNM not applicable
non-Hodgkin's lymphomas	TNM not applicable
Note: involvement of lymph nodes that are not listed as "regional" must be considered as distant metastasis (M1/M+)	

Table 8. Factor C is a measure of the level of certainty of the classification based on the diagnostic methods used BACK

Factor C	Definition
C1	evidence from standard diagnostic means (inspection, palpation, standard radiography, intraluminal endoscopy)
C2	evidence obtained by special diagnostic means (radiographic imaging in special projections, tomography, ultrasonography, lymphography, angiography, scintigraphy, MRI, endoscopy, biopsy, and cytology)
C3	evidence from surgical exploration, including biopsy and cytology
C4	evidence of the extent of disease following definitive surgery and pathological examination of the resected specimen
C5	evidence from autopsy

Table 9. Classification of patients based on screening BACK

Status	Description	Notes
1	cancers screen-detected at first screening test	i.e., cases diagnosed among subjects called back after positive mammogram/Pap smear/FOBT at their first test in the program (status = 10 if they tested spontaneously)
2	cancers screen-detected at a screening test subsequent to the first*	i.e., cases diagnosed among subjects called back after positive mammogram/Pap smear/FOBT, with negative outcome of the previous test taken within the local screening program (status = 20 if they tested spontaneously)
3	cancers identified during the screening interval*	i.e., cases diagnosed after the last screening test with a negative outcome (status = 30 if they tested spontaneously)
4	early recall cancers	early recall cancers are defined as follows:
41	within 6 months from test	❖ cases with previous negative screening test, that were not referred for further examination but invited to repeat the test
42	6 months - 1 year from test	within the intervals described (41-42-43)
43	1-2 years from test	❖ cases with previous positive screening test, referred for further examination and invited
44	2-3 years from test	at one or more subsequent follow-ups (41-42-43); if they presented spontaneously, the status is the same number multiplied by 10 (40; 410; 420; 430; 440)
5	cancer arising in subjects who explicitly refused the invitation	i.e., case diagnosed in subjects who were not invited or not invited again at their own request
6	pre-*screening cancers*	i.e., cases diagnosed before their first invitation, in subjects not yet invited but belonging to the target groups; this group can include cases resident in areas (or townships) where the Local Health Unit (ASL) began screening late, cases arising in subjects who just entered the age class involved in screening and not yet invited, or cases arising in subjects who just migrated to the towns involved in the screening and not yet invited
7	cancer arising in subjects excluded from screening	i.e., cases diagnosed in subjects excluded from screening for specific reasons, which must be stated, where possible, using subclasses (71-72-73)
71	age	
72	specific reason (program-based)	
73	error	
8	cancer arising in subjects who did not respond to the invitation	i.e., cases diagnosed in subjects whose characteristics must be stated, where possible, using subclasses (81-82)
81	responded to at least one invitation, did not attend last screening test	
82	never responded to screening invitation	
9	screening status dubious or unclassifiable	i.e., cases diagnosed in subjects that do not fall into the previous definitions or for whom recorded data is not sufficient to classify their screening status with certainty

**fill in the date of the last screening test*