Chapter 2. Editorial process

Eva Steliarova-Foucher, Murielle Colombet, Anastasia Dolya, Lynn A. G. Ries, Florencia Moreno, Eric Masuyer, Jérôme Vignat, Hee Young Shin, Peter Hesseling, and Charles A. Stiller

CALL FOR DATA

The IICC-3 study was coordinated by IARC. Potential contributors were identified in the IACR membership list, through relevant publications, and through personal communication. They were invited in a personalized letter to contribute to the study. The content of the call for data was also released on a dedicated website (https://iicc.iarc.who.int/).

The target period for IICC-3 included all calendar years starting with 1990, until the last complete year available in a registry by 2013 or 2014. If a registry had participated in previous volumes of IICC, the first required calendar year was to immediately follow the last year contributed to a previous volume whenever possible. The target age range was 0–19 years, but it was 0–14 years for the paediatric cancer registries covering a population younger than 15 years. All primary malignancies, together with non-malignant primary tumours of the CNS, were eligible for inclusion.

Prospective contributors were asked to supply three sets of documents: (1) cancer case records, (2) population data, and (3) a registry profile.

Cancer case records

The file of cancer data consisted of one record for each cancer, containing coded information on the 12 variables described below. Most of these variables were mandatory, and their absence resulted in the exclusion of a given record from the dataset.

Patient identification number

Any format of identification number was accepted. The only requirement for the identification number was the possibility of a linkage of the submitted and registered information on each case by the registry, to enable verification if a query was generated at IARC. This was a mandatory variable.

Tumour sequence number

For this mandatory variable, the recommended coding was as follows:

0 A single tumour in an individual
1 First tumour of two or more in an individual
2 Second tumour in an individual
3 Third tumour in an individual
etc.

Using the combination of the patient identification number and the tumour sequence number, all neoplasms diagnosed in a single individual could be grouped and the occurrence of multiple primaries could be verified.

Date of birth and date of incidence

Coding was required for date of birth and date of incidence in the YYYYMMDD format of the Gregorian calendar, where YYYY is the four-digit calendar year, MM is the two-digit month, and DD is the two-digit day. Year of incidence was a mandatory data item.

Age

Age was defined as the last completed year of age reached by the patient at the date of incidence. Age was a mandatory item unless complete date of birth and date of incidence were provided.

Sex

For this mandatory variable, three values were defined: female, male, and unknown.

Ethnic group

This variable denotes any substantial subpopulation within the registration area, defined by race, ethnicity, religion, or nationality. This optional item was only relevant when a registry could categorize all the cases and the population data in the same way.

Topography, morphology, and behaviour

These three mandatory variables describe the primary neoplasm, and the requested coding scheme was ICD-O-3 [1]. If registries were not already using ICD-O-3, they were encouraged to convert their data before submission.

Grade of diagnosis

This optional variable was to be coded according to ICD-O-3 [1].

Basis of diagnosis

The most valid basis of diagnosis was a mandatory variable, which conveys the level of certainty of the reported diagnosis. The recommended coding scheme was ICD-O-3; however, for the purposes of IICC-3, all codes were converted into four values: microscopically verified (MV), clinical observation (including imaging and biochemistry), diagnosis on the basis of death certificate only (DCO), and unknown.

Laterality

Laterality was requested as an optional variable for the tumours occurring in certain paired organs, i.e. the eyes, kidneys, and gonads. Registries were asked to distinguish between unilateral and bilateral tumours.

Population data
The contributors were asked to supply information on the composition of the population resident in the area from which the cases were drawn. The mid-year population estimate was requested for each calendar year of the covered period, separately by sex and single year of age between 0 and 14 years or between 0 and 19 years for the total covered population and for each defined subpopulation. The sources of the population data and any methods of data production other than those provided by the official statistical institutions were also to be reported.

Registry profile
In addition to data on cancer cases and population data, each registry was requested to supply information describing its registration practices and context in predefined ways.

Questionnaire
In a questionnaire, the registries were asked about the covered area and population, starting year of data collection, definition of reportable cases, coding practices, data sources, and information collected on each case. The source of population data was to be cited. The existence of any screening programmes during the study period was to be reported. Updated contact details and a formal agreement to participate in the study were also included.

Narrative
A short text (with a maximum of 500 words) was requested, describing the registration area, the nature of medical services available for children with cancer, the data sources, and the registry background and operations, according to a standard outline. Specific notes on the presented data and one or two references for further reading could be included.

Map
Each registry covering a subnational population was required to provide a map locating its registration area within the country.

DATA SUBMISSION AND PROCESSING
All contributions were submitted in electronic format through the Registries Portal at https://cinportal.iarc.fr. Connection to the portal required a registry-specific username and password. The username served as an automatic pointer to the registry-specific folder on an internal secured server and constituted a part of the renamed filename of each submitted item. In addition, the registries were asked to specify the type of information that each uploaded file contained, such as cancer cases, population, introductory text, or some others. Subsequently, the filename of each uploaded file was completed with a relevant extension describing the file contents and the date of submission.

The submissions were coordinated by national networks in several countries. Some of these networks were contributors themselves and were therefore counted as contributing registries (NPCR and SEER), whereas others acted as facilitators (Chile, China, Japan, France, Italy, and Poland). The network submissions had the advantage of easing the burden on the individual registries and encouraging participation of the constituent registries, although the follow-up communication was often conducted directly with the individual registries.

The data flow is schematically illustrated in Fig. 2.1.

Cancer data
After the cancer data were received, each file of the cancer records was scanned for the mandatory variables, the eligibility criteria, the coding systems used, and the format. New variables were created for the purposes of data checks or analyses. The age was calculated from the provided date of birth and date of incidence and compared with the provided age at incidence. If the tumour sequence number was missing or inconsistent, it was created using the patient identification number, sex, and dates of birth and incidence. Any discrepancies were discussed with the submitting registry, and the final values were confirmed by the registry.

The next step included conversions into standard coding schemes if required. Topography, morphology, and behaviour coded according to other systems were converted to ICD-O-3 [1] using IARCcrgTools, version 2.10 [2]. The records that could not be converted were listed and sent back to the registry for validation or correction. This list included the cases with behaviour coded to 6 (metastasis) or 9 (unknown whether primary site or metastasis), which were excluded from the analyses if they were not resolved. More information on the coding systems used and any conversion that was done centrally for each registry can be found in the narratives, in the Editors' comments, and in Table A.10.

The records with complete ICD-O-3 codes were then converted according to ICCC-3 [3], modified as described in Chapter 3. The conversion code was written in Stata/IC 12.1 [4].

At the next stage, the dataset was run through specific routines developed at IARC for this purpose in Stata/IC 12.1 [5] and R-3.3.1 [6], based on international standards [2] and additional checks, described in Chapter 4. These routines also produced editorial material and feedback files for the registries. The feedback files contained lists of excluded records and of queried records.

Types of excluded records:
- Records with any mandatory variable missing.
- Multiple records with identical combination of the patient identification number and the tumour sequence number.
- Records that it was not possible to classify into the 2017 update of ICCC-3 (ICCC-3-2017), described in Chapter 3.
- Records considered as multiple primaries according to the IARC/IACR rules, updated in accordance with the European Network of Cancer Registries recommendations on coding haematological malignancies [6].
- Second or further subsequent records of unilateral retinoblastoma, nephroblastoma, or gonadal tumour in one patient; the laterality of the remaining record was recoded to bilateral in the IICC-3 database.

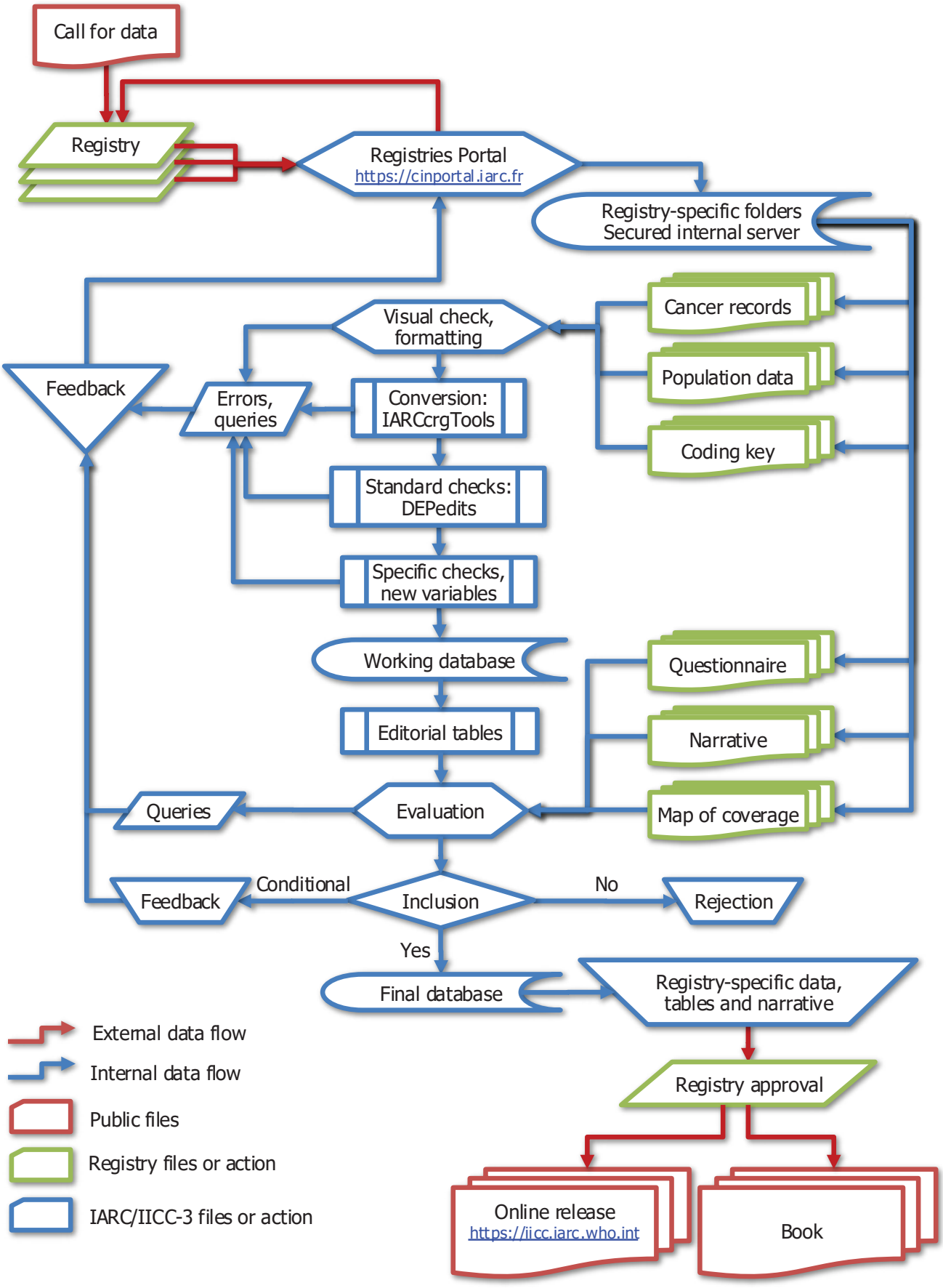

Fig. 2.1. IICC-3 data flow and editorial process.

Types of queried records:

- Provided age at incidence different from the age calculated from the provided date of birth and date of incidence.
- Tumour sequence number indicating:
  - more than one tumour in a patient, with an absence of the matching record(s) of the other primary tumour(s) in the same patient in the dataset;
  - only one tumour in a patient, whereas several were found in the dataset;
  - a missing value.
- Unknown basis of diagnosis.
- Unlikely combination of histology, site, or both with age, according to IARCcrgTools [2].
- Unlikely combination of site with histology, using the morphological families described in IARCcrgTools [2] and updated with new morphology codes described in Chapter 3.
- Unlikely combination of histology with basis of diagnosis, assuming that the specific morphology codes must be supported by diagnosis using microscopic verification, with described exceptions [1].
- Unspecified morphology codes, as defined in Chapter 4.
- Carcinoma in the brain (other than those allowed by ICCC-3-2017), as an implausible combination.
- Neuroblastoma in the brain, as a very rare entity that may be a metastasis.
- Carcinoid, as a rare morphology before age 20 years with a wide variation in occurrence because of the local practices in coding the behaviour of this tumour.
- Carcinoma of the appendix, colon, or small intestine, as rare entities in the target age range.
- Other and unspecified carcinomas (XIf diagnostic subgroup of ICCC-3-2017) in infants, which almost never occur and may be a result of miscoding of age or histology.
- Rare tumours of the bone and soft tissues.

The above-mentioned queries were based on definitions described in Chapter 4.

Population data

The submitted population data file was first checked for the available details. If a population count was not provided for any combination of calendar year, sex, single year of age, and ethnicity, additional official data sources accessible from IARC to complement the submitted data were sought. Subsequently, data were estimated for any missing combinations of categories, as described in the "Statistical methods" section below.

When the data were complete for all relevant combinations, they were examined for consistency in tables and graphs showing the annual population by calendar year, sex, and age group and the evolution of counts across single years of age for each sex, separately in each ethnic group. Unusual patterns were discussed among the editors and with the registries. This may have resulted in a resubmission of population data, additional explanations (e.g. a change in the registration area), or a restriction of the period included in IICC-3. Thus-processed population data were used for the calculation of incidence rates.

IICC-3 database

The consolidated list of cancer records of a registry was appended to a working IICC-3 database, which was designed in Microsoft Structured Query Language (SQL) server 2000 (https://msdn.microsoft.com/en-us/library/ee796613(v=cs.20).aspx). The file containing the detailed split of population counts was also added to the database. Several registry-specific tags characterizing a given registry were also added. They included information on the eligible age range, the instructions for using the provided age or the calculated age, and any ethnic subdivisions. The content of the working database was regularly updated with corrected or additional datasets.

The final version of the registry cancer data and population data that were accepted for publication constituted the final IICC3 database. The final database was frozen and stored on an internal server, and it was used to produce the output for the book and the online resources. It will also be used in ensuing approved specific projects in collaboration with the contributing registries.

Questionnaire

The online questionnaire was developed on classic Active Server Pages (ASP 3.0, released in November 2000 as part of IIS 5.0; https://support.microsoft.com/en-us/help/2669020/active-server-pages-asp-support-in-windows). The contributors could access the questionnaire through the Registries Portal, and it remained open for updates throughout the period of the study. A copy of each version of the questionnaire was sent automatically to the registry and recorded at IARC in PDF format. The responses were stored in an SQL database (Microsoft SQL Server 2000; https://msdn.microsoft.com/en-us/library/ee796613(v=cs.20).aspx). The completed questionnaires were reviewed and assessed for consistency with the other material received, and selected information is tabulated in several Annex tables.

STATISTICAL METHODS

Population at risk

The population data from each registry were arranged into a matrix containing a population count for each combination of calendar year of the study period, sex, and single year of age. A separate matrix was constructed for each subpopulation defined by ethnic group or other group. Data for missing combinations were estimated using all other available details. In instances of conflicting figures, the priority for use of data source followed a defined order. Population data from a census were considered to be the most reliable, followed by official estimates, followed by other estimates. The extent of detail of the population data provided and estimates made centrally for each registry was reported in the "Population at risk" section of the registry narratives, and it was summarized in Table A.8.

Missing counts for calendar years

If population counts were missing for some of the calendar years of the study period, these were estimated by linear interpolation of trends between two years with known counts. The linear interpolant is a straight line between two known points, each defined by their coordinates x_i and y_i. If the two known points have the coordinates

(x_1, y_1) and (x_2, y_2), then the y_i value for a point x_i is $y_i = y_1 + (x_i - x_1) \times [(y_2 - y_1)/(x_2 - x_1)]$.

If calendar years with provided population data lay within, but did not include, the first or last year(s) of the IICC-3 period, the data for such missing years were estimated by linear extrapolation as described above, using the coordinates of the farthest points on the trend line.

The above-mentioned linear trends were derived within the relevant subcategories (e.g. sex and age group), and the resulting estimates were summed across the categories to obtain the totals for the missing years. A known population growth rate was used in the absence of other relevant data.

Missing counts for sex or age strata

Sometimes the population data for the registration area were not split by sex or age. In such instances, the closest possible data source was sought to provide information on the representation of each sex and age category within a total. If the national data were split by sex and age, their sex- and age-specific proportions were applied to the regional data. If sex- and age-specific counts were provided for only some calendar years, their proportions were applied to the other calendar years, either as mean proportions for the intermediate years or as the proportions in years closest to the years at the extremes of the available period. If the sex ratio was provided only in official figures, it was used to calculate the counts by sex. If only totals for all ages were provided, published proportions of 5-year age groups within that population were used to derive the counts in the age groups. The sex- and age-specific proportions were also used to estimate population counts for some or all of the calendar years of the study period.

Missing counts for individual years of age

The counts for single years of age were the most frequently missing information. Usually, data were provided for the 5-year age groups (and exceptionally for the age groups 0, 1–4, 5–9, 10–14, and 15–19 years) instead of for single years of age. If the counts for single years of age were provided for some calendar years or for some overarching population, their proportions were used to calculate the counts for the missing single years of age. Often, the proportions of the population in individual years of age were derived from the preceding or succeeding calendar years with the relevant data or extracted from the data for census years. Alternatively, the United Nations estimates (available from https://unstats.un.org/unsd/demographic/products/dyb/dybcens.htm) were used. If no details were available, the population counts for age groups were split equally to obtain the population counts for single years of age for each sex and calendar year.

Missing counts for categories within subpopulations

The year, age, or sex distribution of the population within the subpopulations of a registration area was not always available, and the estimates of the missing figures were based on the same principles as described above. In the absence of relevant information for a given ethnic group, the year, sex, and age composition of the overall population was applied to the totals for the ethnic group.

Implications of the estimations

The above-mentioned methods were based on various assumptions, which may have been more or less valid. Usually, a stable population growth for the total population or within the subcategories and an identical age–sex constitution over time, across a country, or within subpopulations was assumed. In general, the precision of the estimated population counts decreased with distance from the known data points and with increasing number of data points that needed to be estimated.

Population data tables

The population data used to calculate incidence rates are tabulated along with the standard incidence tables for each dataset in the book or on a page that follows the standard incidence tables displayed online.

In population tables, the average annual population is the result of the division of the total number of person-years by the number of calendar years covered.

The population at risk for combined datasets was constructed as the sum of all person-years contributed by all the registration areas within the period, sex, and age categories. The period reported in the tables for each such pool includes the first and last calendar years covered by any of the contributing registries. For combined datasets, the average annual population was only reported when all the constituent registries covered the same time period. Similarly, if the definition of the registration area changed during the IICC-3 period, the average annual population was not reported. Any change in the registration area is described in the narrative for the relevant registry in Chapter 6.

Percentage

Percentages are reported in several types of tables. In the standard incidence tables, the percentage of cases within the total or within a main diagnostic group is calculated as follows:

Percentage All = $(n_{ij}/N) \times 100$

Percentage Group = $(n_{ij}/n_i) \times 100$.

In the formulas above, n_{ij} is the number of cases in a diagnostic subgroup *ij*, n_i is the number of cases in the main diagnostic group *i* containing subgroup *ij*, and *N* is the total number of cases in the dataset. For each percentage, n_{ij}, n_i, and *N* refer to the same age range and sex.

Unless otherwise specified, the percentages of microscopically verified cases (%MV), cases registered on the basis of death certificate only (%DCO), and tumours with unspecified morphology or topography (%NOS; "not otherwise specified") reported in various tables refer to the accompanying total number of cases.

Sex ratio

The sex ratio is provided in comparative tables for some diagnostic entities in this book and in the online version of the standard incidence tables. The male-to-female sex ratio (M/F) is calculated as n_m/n_f, where n_m is the number of males and n_f is the number of females in the same group of cases defined by the row heading in a given table. In Table A.9, the sex ratio is calculated from the age-standardized incidence rate (ASR) as ASR_m/ASR_f, where ASR_m is the ASR for males and ASR_f is the ASR for females, and therefore compares the incidence rate in males with that in females.

Incidence rates

The incidence rate λ expresses the frequency of cancer in a defined population over a defined time period. It is calculated by dividing the number of cases *n* characterized by a given period, sex, age, diagnostic group, and so on by the person-years *PY* that give rise to those cancers. Thus, both the cases and the person-years used in the formula refer to the same calendar period, the same sex, the same age group, and the same ethnic group. To avoid reporting very small numbers, the resulting number is multiplied by 1 million, so that the incidence rates reported in this publication are always expressed per million person-years, according to the formula

$\lambda = (n/PY) \times 1\,000\,000$.

Age-specific incidence rates refer to the incidence within a specified age group.

The ASRs are calculated by direct standardization as a weighted average of the age-specific incidence rates for 5-year age groups, where the weights are derived from the world standard population [7] within the 5-year age groups younger than 20 years. The world standard population and the notation for the age-specific rates are shown in Table 2.1. Using the notation of Table 2.1, the ASRs were calculated for the age ranges 0–14 years (ASR_{0-14}) and 0–19 years (ASR_{0-19}) as follows:

$ASR_{0-14} = [(r_1 \times 12) + (r_2 \times 10) + (r_3 \times 9)] / 31$

$ASR_{0-19} = [(r_1 \times 12) + (r_2 \times 10) + (r_3 \times 9) + (r_4 \times 9)] / 40$

The formula for the age range 0–14 years differs slightly from that used in IICC-2, in which weights of 2.4 for age 0 years and 9.6 for ages 1–4 years were used to calculate the ASR for ages 0–14 years [8].

Table 2.1. Composition of the world standard population for age groups younger than 20 years [7] used in the calculation of standardized incidence rates in IICC-3

Age (years)	World standard population	Age-specific rate per million
0–4	12 000	r_1
5–9	10 000	r_2
10–14	9 000	r_3
15–19	9 000	r_4
0–14	31 000	
0–19	40 000	

Source: Compiled from Segi et al. (1960) [7].

The cumulative rate [9] is the sum of the age-specific incidence rates in all 5-year age groups, where each age-specific rate is given the same weight, corresponding to the number of years in each age group. Thus, using the notation of Table 2.1, the cumulative rate is calculated as follows:

$Cumulative_{0-14} = (r_1 \times 5) + (r_2 \times 5) + (r_3 \times 5)$

$Cumulative_{0-19} = (r_1 \times 5) + (r_2 \times 5) + (r_3 \times 5) + (r_4 \times 5)$

The cumulative rate approximates the cumulative risk of a person developing a cancer of a particular type before the upper age limit (15 years or 20 years in the respective formulas above). Because the age-specific rates are expressed per million, so are the cumulative rates.

EVALUATION OF THE DATASETS

The data processed at IARC were used to generate (1) editorial material for the IICC-3 editors, (2) feedback files for the contributors, and (3) tables of comparable datasets for dissemination.

Editorial material

Registry-specific editorial material had a standard structure, but the extent of information compiled for the editors may have differed considerably. On average this material contained about 50 pages for each individual dataset and included a standard set of tables and graphs, the questionnaire, the narrative, the map, and a variable amount of additional ad hoc information provided by the registry on request. Separate editorial material was prepared for each ethnic group defined and for the combined datasets of registries operating in a single country. Each dataset was discussed by the editors according to a specific standard form developed and completed for this purpose. A decision was made about whether to include a given dataset in the study or to request additional information from the registry to address specific editors' comments.

Occasionally, the editors may have suggested an extension or a reduction of the period covered by the submitted data, a restriction of the covered area to a smaller region or to a specific subpopulation, a review of a specific subset of data, and so on. The underlying goal was to include as many datasets as possible over the longest available time period, so that a maximum temporal, geographical, and ethnic variety is represented while ensuring the best quality indicators.

Feedback for the contributors

On the basis of the editors' evaluation, each registry received feedback detailing specific features of the dataset, suggestions for further correction or verification of the data, and ad hoc questions to clarify specific patterns. Typically, the feedback fuelled a resubmission of modified files and responses to queries, leading in turn to a new round of data validation until a decision about inclusion or exclusion of the dataset was reached, sometimes despite insufficient responses or no response. This communication to the registries consisted of several elements, as described below.

Follow-up letter

A standardized letter informed the registry about the evaluation of the dataset and the registry status with respect to contribution to IICC-3. The letter contained a description of the submission route, the contents of the files received, any desirable additions to the submission, comments on the observed patterns of incidence and an invitation for their interpretation, a description of missing data items, and a structured account of the data quality indicators. The registries may also have been asked complementary questions on coverage, registration criteria, data sources, actions taken to ensure data completeness and quality, assessment of ethnic status, links with other cancer registries, and population data.

Summary of results
The results were presented for each subpopulation separately and included charts showing the annual distribution of numbers of cases by sex, age group, and diagnostic group; the age-specific distribution of 18 selected tumour types; the proportional distribution of cases according to the four values of the basis of diagnosis by calendar year and age; the distribution of CNS tumours by behaviour, calendar year, and age; the distribution of laterality by calendar year and age for the selected tumour types; the annual distribution of population counts by sex and age group; the age-specific distribution of population counts by sex; and the annual incidence rates by sex and age, overall and for selected diagnoses. A set of six tables laid out similarly to the standard incidence tables presented in this volume and showing the overall numbers, proportions, incidence rates, and quality indicators for all diagnostic categories defined in ICCC-3 for both sexes combined and separately completed this account. This standard output was presented on 20 pages for each dataset.

Files of selected records
Several files of cancer records were prepared, each addressing a different type of query and targeting a specific item of a record; therefore, some records may have been included in several feedback files. Standard naming of the feedback files helped to identify the type of checks required to correct the data, and all queries were described in the follow-up letter. The files were made available for download from the Feedback page on the Registries Portal, and the registry was notified automatically when the file became available. Errors or queries were grouped as follows.

1. Conversion queries
Conversion queries were the records that required validation after a conversion from the coding system used in the submitted data (e.g. ICD-10 site code + ICD-O-2 morphology code) into the ICD-O-3 codes (morphology and topography). Many of these records were haematological malignancies, and they were excluded from the analyses unless they were corrected in a resubmitted file.

2. Ineligible records
Ineligible records were excluded from the analyses because of a problem with one or more mandatory variables. The principal problems, identified in the error file, comprised topography coded to an ICD-10 code that does not correspond to any ICD-O-3 topography code, missing histology code, missing behaviour code, and occasional registrations from years outside the study period or age range, among others.

3. Unclassified records
Unclassified records were not included in the analyses because they could not be classified according to ICCC-3-2017 (described in Chapter 3). Typically, this file included non-malignant tumours occurring outside the CNS, tumours with behaviour code > 3, or records with coding errors. The registries were also provided with hints for possible recoding of the unclassifiable records in a detailed document called “ICCC-3 unclassifiable records”, placed at the disposal of the registries on the Registries Portal.

4. Queried records
Queried records were included in the analyses but contained items that needed to be confirmed or modified. Many of the listed problems related to systematic coding habits. The registries were supplied via the Registries Portal with a complete guide to “IICC-3 queried records”, which suggested the amendments.

5. Duplicate records
Sets of duplicate records with identical combination of the patient identification number and the tumour sequence number were excluded from the analyses until they were corrected and resubmitted.

6. Missing laterality
Records with missing laterality information were listed for all the relevant cancer cases for which laterality was requested but was unknown, so that the registries could focus on obtaining the missing information for the listed records if possible.

7. False multiple primaries
Sets of records pertaining to neoplasms found in the same patient, considered as a single tumour according to the IARC/IACR rules for multiple primaries, were listed as false multiple primaries. Each set of such records was replaced with a consolidated record, combining the best information from the other records and included in the analyses. The other records of the same set were excluded. If any of the listed sets were cases of retinoblastoma, nephroblastoma, or gonadal tumours, the laterality may have changed in the consolidated record, so that two unilateral cases were recorded as a single bilateral case. The registry was invited to modify the original records as appropriate and resubmit a corrected dataset.

Selection of datasets
The two guiding principles for the selection of the datasets were maximum data availability and maximum data comparability. Both principles needed to be satisfied for a dataset to be selected, although the availability principle prevailed when deciding about the inclusion of registries operating in world regions with a dearth of cancer data. The data quality criteria may have been relaxed slightly with respect to the proportion of MV cases, unconfirmed underreporting or overreporting, lack of specific details (such as precision of the dates), and so on, because they were conditioned by dynamics external to the registry.

Many registries demonstrated keen cooperation and improved their data considerably. There were 125 datasets that were not considered acceptable on first submission and were included in IICC-3 by the end of the editorial process (Table 2.2). The implemented improvements included completion of missing information, adding more recent or earlier years of data, attending to missed data sources, correcting coding errors, improving age calculations, including non-malignant CNS tumours, replacing population data, and explaining queried patterns in the data. The quality of a response and the interpretation of the observed shortcomings by a registry constituted an important element in the evaluation process.

Table 2.2. Overview of the participation of registries in IICC-3

Continent	Invited (N)	Submitting (N)	Total evaluations (N)	Quality improved (N)	Included (N)	Initially not included (N)	Success rate* (%)
Africa	70	29	100	17	21	15	72.4
America, Latin and the Caribbean	91	54	169	26	38	17	70.4
America, North	85	78	242	7	70	2	89.7
Asia	157	103	270	44	58	25	56.3
Europe	214	149	438	70	109	63	73.2
Oceania	19	12	20	3	12	3	100.0
Total	**636**	**425**	**1239**	**165**	**308**	**125**	**72.5**

N, number of registries.
*Success rate is defined as the ratio of Included to Submitting.

The most common reasons for rejection included large fluctuation in numbers of cases or in annual incidence rates over time, rates that were too high or too low (overall or by diagnostic group), many unlikely cases for the age at diagnosis, an unlikely case mix, a large proportion of DCO cases, a low proportion of MV cases, a large proportion of cases with unspecified morphology or topography (NOS), too few cases to provide a reasonably stable national estimate of incidence, and lack of response about the notified shortcomings. Multiple deficiencies were almost always the basis for a rejection.

The invitations to join the study were sent to more than 600 registries, of which more than 400 submitted their data. The submitted material generated more than 800 datasets (accounting for subpopulations and combined datasets), which were evaluated separately in more than 1200 instances. However, the painstaking editorial process bore fruit, with data quality improved in 165 registries (39% of the submitting registries), including 125 of those included in IICC-3 (40% of 308). On average, 72.5% of the submitting registries were included, but this proportion differed across continents (Table 2.2). The significant improvement in data quality testifies to the importance of participation in international studies for the improvement of local data. Improved data quality benefits not only the IICC-3 study but also all studies conducted using the childhood cancer data, whether locally or internationally.

DESIGN OF THE OUTPUT

The content, format, and layout of the study results were inspired by other international projects, such as previous IICC volumes and the *Cancer Incidence in Five Continents* series (e.g. [10]), but IICC-3 also includes certain novel features, notably a dual publication (book and online versions) and novel comparative tables presenting input and output data.

Because of the large number of datasets included in IICC-3, it proved impractical to present in a book the results for many more registries with the same detail as was done in the two previous volumes. A selection was required to present the maximum available data in the limited space offered by a book to be printed. In response, combined national datasets were constructed for many participating countries, and these were selected for display in the book. In countries where the contributors included paediatric registries or the data could be presented by ethnic group, additional datasets were included in the book. However, all tables produced for all eligible datasets were made available online. The availability of all registry-specific tables is mapped in Table A.12.

Tables

IICC-3 contains a large number and variety of tables. The standard registry-specific incidence tables (Chapter 6) and the comparative tables showing the data for all eligible datasets (Chapter 7 to Chapter 13) were developed using Stata/IC 12.1 [4] in association with MiKTeX (https://miktex.org) and complemented by LaTeX (https://www.latex-project.org). The other tables, notably the Annex tables, were constructed in Microsoft Excel.

Narratives

The editors reviewed the registry narratives for content, structure, and style and edited them to ensure as much coherence and standardized content as possible. For the countries represented by a combined dataset, the editors also prepared a country overview as a brief introduction, describing the country location, the demographics, the health-care system, and the combined dataset produced for IICC-3. Any common features with respect to data sources or data quality were also integrated into this introductory part of the narrative and omitted in the registry-specific sections. The "Population at risk" section describes the data sources and the detail with which population data were provided. The resulting registry-specific narrative section is based on the submitted text, the completed questionnaire, the actual data received, and additional information from the registry or other publicly available sources accessed by IARC. The content of the narratives was approved by the listed contributors.

Editors' comments

The Editors' comments draw attention to the features in the data that may affect data completeness and comparability, as an aid to interpreting the results. Unlike in the two previous volumes, asterisks were not assigned to the datasets with perceived reduced comparability. Such a binary index was considered to be of limited usefulness for data interpretation and garnered little attention in derived studies. Therefore, the "Editors' comments" section should be regarded as an integral part of the presented results and should guide the selection of registries for ensuing detailed studies. These comments are complemented by the extensive Annex tables, in which summary statistics may be compared across all registries.

Maps

The maps indicating the location and coverage of the submitting registries were checked for consistency with the questionnaire, the narrative, and the data and then used as a model for developing the maps displayed in this publication. The colour scheme distinguishes the type of contributing registry (general or paediatric), and the coloured area enables recognition of national or regional coverage. The R language and environment was used to generate the maps [5], using the ggplot2 data visualization package for R (https://www.r-project.org). The ggplot2 package facilitated the creation of maps by superposing multiple layers from different data sources and automatically adjusting common scales [11]. Shape files and map templates were based on the standard operating procedures for the creation of maps by WHO standards. The maps were further improved with Inkscape software (https://inkscape.org), which enabled splitting the frames and adding titles and legends.

Online resources

The development of the online resources at https://iicc.iarc.who.int/ required specific attention to the content, design, and technical details. The online resources offer a large amount of additional data, which was arranged for intuitive access. A specific IICC logo was designed to accompany the online material and brand its source. The page displaying the study results was developed using hypertext preprocessor (PHP; https://secure.php.net). The contents of the online resources are described in Chapter 5.

REFERENCES

1. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al., editors (2000). International Classification of Diseases for Oncology. 3rd ed. (ICD-O-3). Geneva, Switzerland: World Health Organization.
2. Ferlay J, Burkhard C, Whelan S, Parkin DM (2005). Check and conversion programs for cancer registries (IARC/IACR tools for cancer registries). IARC Technical Report No. 42. Lyon, France: International Agency for Research on Cancer.
3. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P (2005). International Classification of Childhood Cancer, third edition. Cancer. 103(7):1457–67. https://doi.org/10.1002/cncr.20910 PMID:15712273
4. StataCorp (2011). Stata statistical software: release 12. College Station (TX), USA: StataCorp LP.
5. R Core Team (2017). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org.
6. Gavin A, Rous B, Marcos-Gragera R, Middleton R, Steliarova-Foucher E, Maynadie M, et al.; European Network of Cancer Registries (2015). Towards optimal clinical and epidemiological registration of haematological malignancies: guidelines for recording progressions, transformations and multiple diagnoses. Eur J Cancer. 51(9):1109–22. https://doi.org/10.1016/j.ejca.2014.02.008 PMID:24630945
7. Segi M, Fujisaku S, Kurihara M, Narai Y, Sasajima K (1960). The age-adjusted death rates for malignant neoplasms in some selected sites in 23 countries in 1954–1955 and their geographical correlation. Tohoku J Exp Med. 72(1):91–103. https://doi.org/10.1620/tjem.72.91 PMID:13749685
8. Parkin DM, Kramárová E (1998). Materials and methods. In: Parkin DM, Kramárová E, Draper GJ, Masuyer E, Michaelis J, Neglia J, et al., editors. International Incidence of Childhood Cancer, Vol. II. IARC Scientific Publication No. 144. Lyon, France: International Agency for Research on Cancer; pp. 3–14.
9. Day N (1982). Cumulative rate and cumulative risk. In: Waterhouse J, Muir C, Shanmugaratnam K, Powell J, editors. Cancer Incidence in Five Continents, Vol. IV. IARC Scientific Publication No. 42. Lyon, France: International Agency for Research on Cancer; pp. 668–70.
10. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, et al., editors (2014). Cancer Incidence in Five Continents, Vol. X. IARC Scientific Publication No. 164. Lyon, France: International Agency for Research on Cancer.
11. Wickham H (2009). ggplot2: elegant graphics for data analysis. New York (NY), USA: Springer-Verlag.