Chapter 4. Data quality and comparability

Eva Steliarova-Foucher, Charles A. Stiller, Murielle Colombet, Anastasia Dolya, Florencia Moreno, and Lynn A. G. Ries

The cancer registries selected for IICC-3 have complied with the inclusion criteria. The reported frequencies, proportions, and rates from these registries are partly dependent on the structures for health-care provision in their respective countries and the legal and administrative status of the registry, which determine the case-finding sources and data collection routines. The availability and quality of national statistical services have an impact on the possibility of case ascertainment from death certificates and the precision of the population estimates. The density and accessibility of health-care facilities influence the timeliness and precision of diagnosis and the extent of underdiagnosis. Social customs and preferences may cause reporting bias for some population subgroups (e.g. infants, girls, ethnic groups) or impede the collection of selected data items (e.g. exact date of birth).

At the registry level, the practices may also differ. The reportable cases may be defined with respect to the behaviour of the neoplasm (e.g. only malignant tumours), tumour type (e.g. exclusion of skin carcinomas), definition of eligible residents (based on origin, place of birth, or length of residence), data items collected (e.g. grade, laterality), the precision of the reported information, and the extent of its validation using multiple data sources.

All these background factors can influence the observed incidence rates. Therefore, this chapter is devoted to a description of the underlying differences that need to be considered when interpreting geographical variations reported in IICC-3.

POPULATIONS COVERED

The annual populations of people younger than 20 years covered by the registration areas included in IICC-3 varied in size by more than 1000-fold, ranging from less than 50 000 (e.g. Neuchâtel, Switzerland) to more than 50 million (USA, NPCR). The larger the population at risk, the more reliable the incidence rates are likely to be, assuming similar data quality. In the registration areas described in IICC-3, the proportion of people younger than 20 years within the total population ranged from 11% (China, Shanghai) to 63% (Cameroon, Yaoundé). The registry-specific proportions of childhood population are compared with the equivalent values for the whole country or territory in Table A.6.

Based on the 2010 United Nations estimates of national populations [1], the contributing registries covered between less than 1% and 100% of the childhood population within the represented countries or territories. Combining data from regional cancer registries into a pooled dataset helped to increase the coverage of the country and the national representativeness of the dataset (Table A.6). The larger the proportion of the national population covered, the more representative the presented incidence rates are of the country, and the same is true of the representativeness of a continent or any geographical entity.

Two or more racial or ethnic subgroups were distinguished within the total population of children covered in some countries (Table 4.1). Differences in cancer occurrence between population subgroups within the same country may indicate differences in cancer susceptibility or in exposure to risk factors, but also in access to health care.

ACCESS TO DATA SOURCES

Population at risk

As a rule, the resident population is counted by national statistical services at regular intervals, usually by means of a population census. For the years between two censuses, the population is estimated taking into account births, deaths, and migration. The population counts are usually split by sex and age and are often provided for constituent administrative units. The age categories may be provided as 5-year groups, as single years of age, or occasionally as a different age grouping. The census data usually provide the most detailed information and may also be considered the most accurate. The availability of accurate population data representing the registration areas is usually associated with other indicators, such as the Human Development Index. The national statistical services customarily also provide official estimates for subnational administrative units covered by cancer registries. Where this is not done, the resident population in the covered areas may be estimated by the registry using official sources (e.g. USA, Hawaii, Hawaiian). Correct population counts, including within subcategories, are as important as the completeness and quality of data on cancer. Population sizes and census years in the countries included in IICC-3 are shown in Table A.7.

Despite the clear guidelines in the data submission instructions, registries were not always able to supply the population data in the required detail (sex and single year of age) for each calendar year to match the years of diagnosis submitted. There was a wide variety in the extent and detail with which the population counts were provided, and this may have varied by calendar years or in population subgroups. The details of the submitted population data and the contribution of estimates to the data used in the analyses are indicated in Table A.8. The “Population at risk” sections in Chapter 6 provide further information on data sources and availability for

Table 4.1. Overview of multiple populations occupying the same geographical unit, included in IICC-3

Country, registry, ethnic group	Population counts in the reference year**			Reference year	Data source for total population***
	Age 0–14 years	Age 15–19 years	Total population		
SOUTH AFRICA, paediatric	15 100 089	5 003 477	51 770 560	2011	2
SOUTH AFRICA, paediatric, Asian	258 602	98 556	1 286 930	2011	2
SOUTH AFRICA, paediatric, Black	12 702 324	4 171 450	41 000 938	2011	2
SOUTH AFRICA, paediatric, Coloured	1 311 811	431 263	4 615 401	2011	2
SOUTH AFRICA, paediatric, White	771 187	284 896	4 586 838	2011	2
USA	61 227 213	22 040 343	308 745 538	2010	3
USA, API	3 422 656	1 173 201	16 993 326	2010	3
USA, Black	10 102 891	3 775 192	42 065 334	2010	3
USA, Hispanic White	12 773 847	4 027 899	45 295 968	2010	3
USA, Native American	1 144 526	393 320	4 263 538	2010	3
USA, White NH	33 783 293	12 670 731	200 127 372	2010	3
USA, Hawaii, Hawaiian	86 418	27 160	290 580	2010	1
ISRAEL	2 071 599	577 400	7 623 507	2010	1
ISRAEL, Jews	1 489 100	414 800	5 752 170	2010	1
ISRAEL, Non-Jews	582 499	162 600	1 871 337	2010	1
KUWAIT	747 579	197 471	3 566 437	2010	1
KUWAIT, Kuwaitis	434 303	120 333	1 133 214	2010	1
KUWAIT, Non-Kuwaitis	313 276	77 138	2 433 223	2010	1
SAUDI ARABIA, Saudi Arabia*	7 188 313	2 630 062	27 236 156	2010	4
SAUDI ARABIA, Saudi Arabia, Saudis*	5 550 117	2 206 355	18 776 510	2010	4
SAUDI ARABIA, Saudi Arabia, Non-Saudis*	1 638 196	423 707	8 459 646	2010	4
SAUDI ARABIA, Riyadh	2 030 288	566 626	6 777 146	2010	4
SAUDI ARABIA, Riyadh, Saudis	1 591 669	472 836	4 296 745	2010	4
SAUDI ARABIA, Riyadh, Non-Saudis	438 619	93 790	2 480 401	2010	4
AUSTRALIA	4 198 698	1 460 048	22 031 750	2010	5
AUSTRALIA, Indigenous*	239 857	70 905	656 735	2010	5
AUSTRALIA, 3 registries*	1 379 418	472 622	6 925 367	2010	5
AUSTRALIA, 3 registries, Indigenous	123 925	35 656	339 687	2010	5
NEW ZEALAND	894 405	322 275	4 367 510	2010	1
NEW ZEALAND, Maori	225 940	67 920	662 840	2010	1
NEW ZEALAND, Pacific peoples	88 850	27 505	279 095	2010	1

API, Asian Pacific Islander; NH, Non-Hispanic.
*Population not described in IICC-3.
**The population counts for ages < 20 years are reported as submitted for IICC-3.
***Source of population data for all ages:
1. Cancer Incidence in Five Continents, Volume XI [10].
2. https://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf
3. https://www.census.gov/data.html
4. https://www.stats.gov.sa/en/
5. https://www.abs.gov.au/

each contributing registry. The lower the quantity and the detail of the provided data, the more estimates were required to complete the missing information and the greater the risk of the estimates not being representative. The most serious errors may result from extrapolation of population for missing years, followed by interpolation for missing years and by splitting the overall annual data into various subgroups. The impact on the incidence rates varies according to the nature of the missing information. For example, missing years of data will influence the annual rates both overall and in sex and age subgroups, as well as the rates for the entire period. Lack of age details may bias the age- and sex-specific incidence rates. The first year of life may be affected most, because

the proportion of infants is higher than the proportion of children in each subsequent year of age in all populations with positive growth and especially in low- and middle-income countries. Therefore, estimating the population of age 0 years as one fifth of the population of the age group 0–4 years may result in overestimation of rates in infants and overall.

Death certificates

The collection of mortality data is outside the scope of cancer registry functions. Death certification is legally required and conducted in most, but not all, countries. The data contained in death certificates may be centralized at a subnational or national level, may be stored on paper or in an electronic format, and may be identifiable or anonymized. Cancer registries may or may not have access to these data. The availability (http://www.who.int/healthinfo/statistics/mortality_rawdata/en) and quality [2] of mortality statistics in the countries represented in IICC-3 are shown in Table A.7.

Cancer registries access information recorded in death certificates with two aims. The first aim is to ascertain cancer cases that had not been registered from other data sources, and the second is to update the vital status of the patients already in the registry. The extent of ascertainment of new cancer cases from death certificates was assessed in IICC-3. Ideally, all causes of deaths recorded at a national level should be reviewed, and any mention of cancer should lead to a record in a cancer registry, whether national or regional. The next best situation is access to regional databases of deaths from all causes. The accessible data must be identifiable to enable matching of the deceased individuals with those in the registry. If the death certificates can only be reviewed for registered patients, the extracted information is useful for their follow-up, and possibly for ascertainment of new primary tumours, but not for identification of unregistered patients. Non-identifiable death certificates are not useful for identification of new cases, although it may still be possible to compare the frequencies of incident cases with the tabulated mortality data to assess the completeness of ascertainment by the registry for cancers with high case fatality, provided that the tumour category is defined in the mortality statistics.

When a cancer is found on a death certificate, it may either be registered as a new case or an existing record can be updated with the date and cause of death. If a new record is established, the registry searches for evidence of this diagnosis in other data sources. If additional information is found, the relevant record is updated with more precise data on diagnosis and date of incidence. However, if no other information is found, the record remains "death certificate only" (DCO) and becomes part of the DCO statistic. A high percentage of DCO cases suggests that the registry failed to register a proportion of cases between the diagnosis and the death. It can therefore be assumed that a similar proportion of cases may have escaped both registration and cancer appearing on a death certificate. Such cases artificially reduce the reported cancer incidence and mortality rates. A high proportion of DCO cases also affects the quality of the diagnosis, which is less accurate for DCO cases than for cancers ascertained from medical records. However, a complete absence of DCO cases from a registry may mean that death certificates do not constitute a source for data ascertainment, and as a result the reported incidence rates may be underestimated. A full account of the role of the DCO statistic in the assessment of data quality and completeness can be found elsewhere [3, 4].

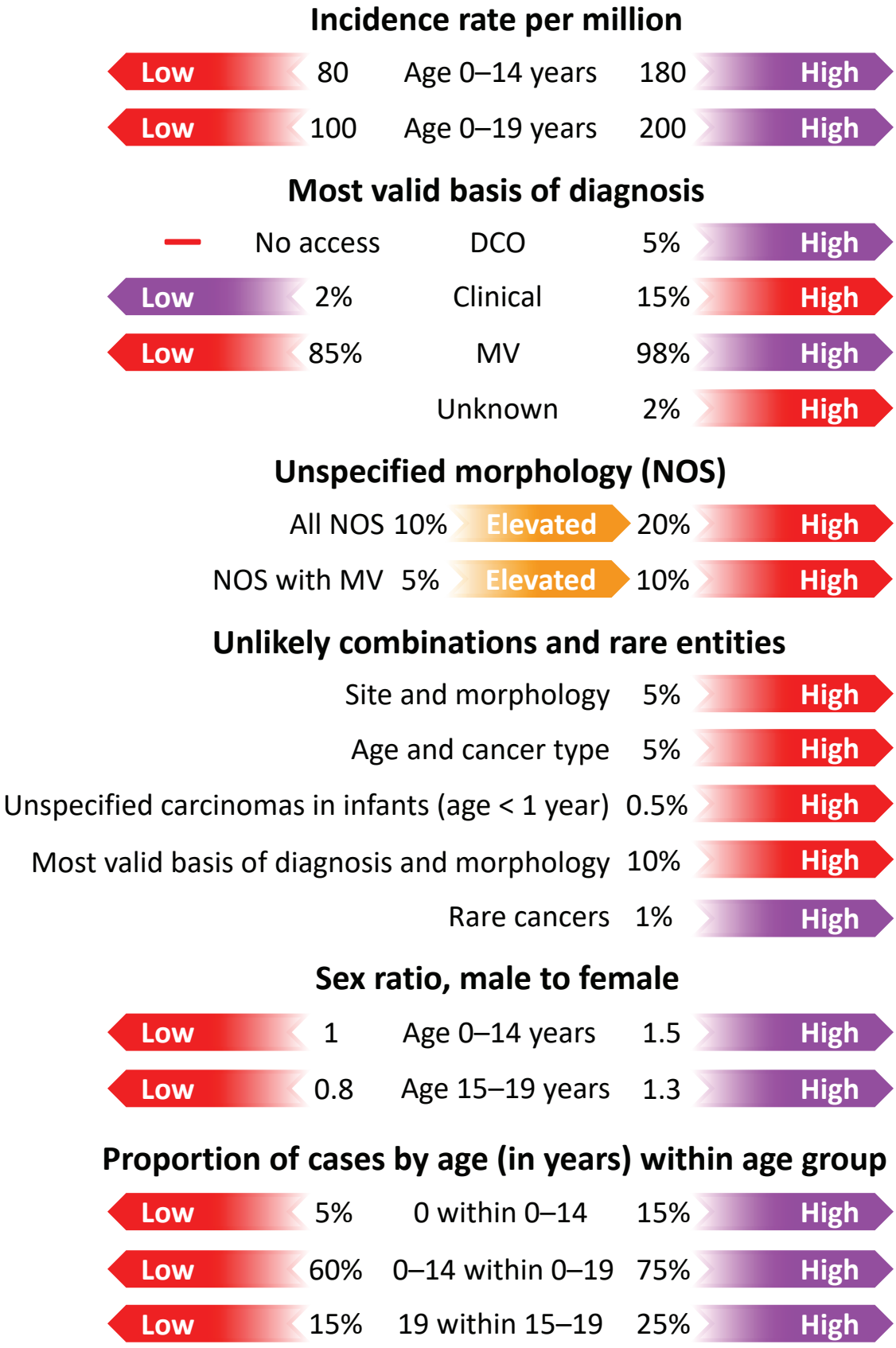
It may be argued that cancer is such a serious condition that it is unlikely for a child with cancer not to have any diagnostic record before death, and therefore that the proportion of DCO cases in children is close to zero. Although this may be true in populations that benefit from a high level of health care and excellent access to diagnostic and therapeutic facilities, a cancer death without a diagnosis, even in a child, is an established fact in many low-income countries. Therefore, the percentage of DCO cases is shown for each diagnostic category in the registry-specific tables. The overall percentage of DCO cases can be compared among the datasets in Table A.9. For the registries that reported no access to death certificates, this item is represented by a red dash in the relevant column. The value 0.0 means either that no cancer cases were identified from examined death certificates or that all identified cases were successfully traced back and additional information was found in medical records. The percentage of DCO cases is highlighted in purple if it is greater than 5% (Fig. 4.1). The percentage was not provided for the pooled datasets, except when all constituent registries reported accessibility to death certificates.

Clinical records

Information on cancer diagnosis is generated by public hospitals, private clinics, paediatric departments, specialized treatment centres, laboratories, and other institutions. In some countries, notification of all cancer diagnoses to the registry is mandatory; in others, the registries mostly visit the relevant institutions and search actively in medical records. Except for DCO cases, all registered cases should be based on at least one medical record. However, registries usually compile each case record from several data sources.

Information on the most valid basis of diagnosis is one of the mandatory data items, and the coding is defined in ICD-O-3 [5]. Based on this standard, in IICC-3 a clinical basis of diagnosis was coded for all cases that were not known to be based on microscopic examination of tumour tissue, except for the DCO cases and those for which the basis of diagnosis is unknown. The proportion of cases defined solely by a clinical diagnosis is relatively low in children with cancer, although some such diagnoses, which may be made unequivocally without tissue examination, do occur in children (Table 4.2).

The overall percentage of cases that are based solely on clinical records can be found in Table A.9 for all datasets that are presented in a separate table. The arbitrary limits of the reasonable range were set to 2–15% and are colour-coded in Table A.9 as indicated in Fig. 4.1. A very low percentage suggests that some cases with only clinical diagnosis may be missed by the registry, whereas a high percentage usually indicates inadequate pathology services or an inability of the registry to access the pathology records.



Note: Certain indicators are not highlighted for specialized paediatric datasets.

Fig. 4.1. Arbitrary cut-off values defining the colour code used for selected indicators displayed in Table A.9 and Table A.11. The listed indicators were not evaluated for the specialized paediatric cancer registries unless otherwise specified. DCO, death certificate only; MV, microscopically verified.

Table 4.2. Neoplasms for which a specific morphology code can be assigned without microscopic examination [9]

ICD-O-3 [5] / ICD-O-3.1 [8]			
Morphology		Topography	
Code	**Term**	**Code**	**Term**
8000	Neoplasm, NOS		
8150–8154	Islet cell tumours, gastrinomas		
8170	Hepatocarcinoma		
8270–8281	Pituitary tumours		
8720	Melanoma of the eye	C69.*	Eye
8720	Melanoma of skin	C44.*	Skin
8800	Sarcoma, NOS		
8960	Nephroblastoma, NOS		
9100	Choriocarcinoma, NOS		
9140	Kaposi sarcoma		
9350	Craniopharyngioma		
9380	Glioma		
9384/1	Subependymal giant cell astrocytoma		
9500	Neuroblastoma, NOS		
9510	Retinoblastoma, NOS		
9530–9539	Meningioma		
9590	Lymphoma, NOS		
9732	Multiple myeloma		
9761	Waldenström macroglobulinemia		
9800	Leukaemia, NOS		

*Applies to the whole range of valid codes that start with the digits shown.
Source: Compiled from Fritz et al. (2000) [5], Fritz et al. (2013) [8], and Ferlay et al. (2005) [9].

Pathology records

Microscopic confirmation of diagnosis

With the current state of knowledge and diagnostic advances, the examination of tumour tissue under a microscope and the ensuing pathology report are considered the most reliable basis of diagnosis, although the pathology assessment is increasingly supplemented by genetic studies. As a result, the high proportion of microscopically verified (MV) diagnoses increases the credibility of the tumour-specific incidence rates. Microscopic verification includes histology (examination of tissue sections from biopsy of the primary tumour or of a metastasis, or of tissue obtained at operation or autopsy) and cytology (haematological examination of peripheral blood specimens).

Although a high proportion of MV cases is desirable, values close to 100% should be regarded with caution because, depending on the registry context, these may suggest that some clinically diagnosed cases were missed. In Table A.9 the percentage of MV cases is shown for all the datasets that are presented in separate tables. Values greater than 98% or less than 85% are highlighted, as shown in Fig. 4.1. The percentage of MV cases is also shown for all the diagnostic categories in the standard registry-specific tables.

Table A.9 displays the percentage of diagnoses with an unknown basis of diagnosis, and values greater than 2% are highlighted (Fig. 4.1). A high percentage of cases with unknown basis of diagnosis reduces the confidence placed on the reported rates.

Unspecified morphology

Another quality indicator, partly related to the proportion of MV cases, is the proportion of cases with unspecified morphology (NOS). For the purposes of IICC-3, all records that are classified in the subgroups of ICCC-3-2017 that gather the neoplasms in the unspecified categories of the main diagnostic groups (Ie, IIe, IIIf, VIc, VIIc, VIIIe, IXe, XIIb), those with unspecified histology codes M-8000 to M-8005 in Xe, and the ill-defined and unspecified sites (C76 to C80.9 only) in the subgroup XIf were included in the overall percentage of NOS cases displayed in Table A.9. The percentage represented by the subgroups of unspecified tumour types can also be assessed in the registry-specific tables. A high proportion of NOS cases indicates a reduced quality of the data and raises

questions about the reliability of the tumour-specific incidence rates; therefore, values greater than 10% of the total cases are highlighted, as indicated in Fig. 4.1.

A high proportion of NOS cases was not always correlated with a low proportion of MV cases. Therefore, in Table A.9 the percentages of NOS cases with microscopic verification are also shown, and high values are highlighted according to the thresholds shown in Fig. 4.1.

A particular subset of the NOS cases includes all those classified in subgroup XIIb with the morphology codes M-8000 to M-8005 and topography codes indicating a possibility of a more appropriate ICCC category; examples are leukaemia (C42), skin carcinoma or melanoma (C44), soft tissue sarcoma or neuroblastoma (C49), retinoblastoma (C69 or C69.2), thyroid carcinoma (C73), neuroblastoma or adrenocortical carcinoma (C74), and lymphoma (C77). Although these cases would still fall within an unspecified subgroup, the overall rates of the main diagnostic groups (e.g. leukaemia, lymphoma, or unspecified) would be more correctly estimated. Following the recoding of some of these cases, to enable them to be reassigned into more specific subgroups, the residual numbers of these tumours are shown in Tables 11.5.a and 11.5.b.

Demographic information

Dates and age

Recording accurate date of birth and date of incidence is necessary to calculate the age at diagnosis with an accuracy of 1 day. Because the spectrum of tumours varies rapidly during early years of age, age should be defined and calculated in the same way in all compared registries. However, as shown by the data on incomplete dates in Table A.8, age may have been marginally biased in some registries. Most importantly, India and some other countries lack customary usage of birth date, because it may not be recalled or reported accurately. Therefore, age is recorded as reported by the family, but no validation of its value is possible in the registry or in IICC-3. In other settings, registries may have decided to collect part of the date only (e.g. month and year), which may have an impact on the exactness of age. Other registries were unable to submit full dates because of data confidentiality considerations. Although these registries may have collected accurate dates and used them to calculate age, it was not possible to review the completeness of the recorded dates and validate the age calculation centrally. In some registries the dates were seemingly complete, but the days or months were not distributed as expected. For example, there may have been an excess of days 1, 15, or 30 in a month or an excess of January or July in a year. Except in small datasets, such excesses almost certainly indicate that the real dates were unknown for some cases and that the registry completed the dates with some agreed or random dummy values. Thus, the percentage of incomplete dates in Table A.8 includes, in addition to the dates with values coded as unknown, also the percentages greater than 5% for a given day in a month and the percentages greater than 10% for a given month in a year.

Whenever both date of birth and date of incidence were provided, the age of the patient was calculated by a standard algorithm with a precision of 1 day for all cases and compared with the provided age. In the absence of complete dates, the exact age could not be confirmed. However, for records with a missing date element the calculated age was required to lie within an adequate interval of the provided age. All records with an inconsistency between the provided age and the calculated age were listed for verification and correction by the registry, and the remaining proportion of such records in the final dataset is shown in Table A.8. Use of the calculated age, as opposed to the provided age, in the analyses is also indicated in Table A.8.

Inexact age recorded by a registry may have affected the overall incidence rates, because one of the selection criteria of the data for the IICC-3 study was the age restriction (cancer incident before age 20 years). By extension, the age-specific incidence rates would also be biased. A particularly strong effect could have been exercised on the incidence rates for the extremes of the age range: age 0 years and age 14 years or 19 years. For example, rates in infants in the registries with incomplete or imprecise dates may be underestimated, as documented in some registry-specific tables and Editors' comments. A high proportion of infants may indicate opportunistic screening for neuroblastoma or early diagnosis of other embryonal tumours, and if this is accompanied by an unlikely case mix, the presence of adult cases with unknown age coded to 0 may be suspected, although such cases should have been eliminated through verification of unlikely combinations of age with tumour (Table 4.3). Extreme values of the proportion of cases aged 19 years in the age group 15–19 years or of the proportion of cases aged 0–14 years in the age group 0–19 years in Table A.9 may also suggest biased age, dates, or other data items. Although such bias was detected and corrected in the course of the editorial process in some registries, an unexpected age pattern is highlighted in Table A.9, using the arbitrary limits defined in Fig. 4.1.

Sex

In the areas where cancer registration is considered complete, the overall male-to-female sex ratio of cancer cases in a population younger than 15 years is about 1.2 [6], i.e. boys with cancer are registered about 20% more frequently than girls. However, in several low- and middle-income countries the proportion of registered boys was consistently higher. A high sex ratio may be explained by a high occurrence of cancer types that are more frequent in boys than in girls, such as lymphomas in the Mediterranean countries. However, where the case mix alone cannot explain the excess of registered boys, a sex-based inequity may be suspected. As mentioned by some registries, sick boys may be more likely to get medical attention than sick girls. Such unequal treatment thus reduces the comparability of the incidence rates, both for the two sexes combined and for the undertreated sex (usually girls). In Table A.9, the sex ratios for the age range 0–14 years and the age range 15–19 years that are outside the limits outlined in Fig. 4.1 are highlighted accordingly. Sex ratios are also reported in the registry-specific tables displayed online and in some comparative tables (see Chapter 5). As noted in Table A.8, sex was not recorded for a very small percentage of cases in a few registries; these cases were excluded from sex-specific analyses.

Table 4.3. Definition of the queried combinations of age with cancer type in IICC-3

ICD-O-3 [5] / ICD-O-3.1 [8]		ICCC-3-2017 (see Chapter 3)	Entity	Unlikely age in years
Morphology	Topography	Subgroup**		
9823		(Ia)	B-cell chronic lymphocytic leukaemia / small cell lymphocytic lymphoma	0–19
		IIa	Hodgkin lymphomas	0–2
9732		(IIb)	Multiple myeloma	0–19
		IVa	Neuroblastoma and ganglioneuroblastoma	10–19
951*		(V)	Retinoblastoma	6–19
		VIa	Nephroblastoma and other non-epithelial renal tumours	10–19
		VIb	Renal carcinomas	0–4
		VIIa	Hepatoblastoma and mesenchymal tumours of the liver	10–19
8970		(VIIa)	Hepatoblastoma	5–19
		VIIb	Hepatic carcinomas	0–4
		VIIIa	Osteosarcomas	0–3
		VIIIb	Chondrosarcomas	0–4
		VIIIc	Ewing tumour and related sarcomas of bone	0–3
8991		(IXa)	Embryonal sarcoma	16–19
		Xb	Malignant extracranial and extragonadal germ cell tumours	5–9
		Xd	Gonadal carcinomas	0–4
9072		(Xa, Xb, Xc)	Polyembryoma	0–19
		XIb	Thyroid carcinoma	0–2
		XIc	Nasopharyngeal carcinomas	0–4
		XIe	Skin carcinomas	0–4
		XIf	Other and unspecified carcinomas	0–2
905*		(XIIa)	Mesothelial neoplasms	0–19
8981		(XIIa)	Embryonal carcinosarcomas	0–19
	C15*		Oesophageal cancers	0–19
<9590	C17*		Cancer of small intestine other than lymphoma	0–19
	C19*		Rectosigmoid junction cancers	0–19
	C20*		Rectum cancers	0–19
	C21*		Anal cancers	0–19
	C23*		Gallbladder cancers	0–19
	C24*		Biliary tract cancers	0–19
	C38.4		Pleural cancers	0–19
	C50*		Breast cancers	0–19
	C53*		Cancer of uterus (cervix)	0–19
	C54*		Cancer of uterus (corpus)	0–19
	C55*		Cancer of uterus (NOS)	0–19
814*	C61*		Adenocarcinoma of prostate	0–19
≠824*	C33*		Cancer of trachea (except for carcinoid tumours)	0–19
≠824*	C34*		Cancer of lung (except for carcinoid tumours)	0–19
≠824*	C18*		Cancer of colon (except for carcinoid tumours)	0–19

*Applies to the whole range of valid codes that start with the digits shown.
**If the code is enclosed in parentheses, only the morphology codes shown are subjected to the check.
<Applies to all numerically inferior morphology codes.
≠Applies to all morphology codes that differ numerically from the values shown.
Source: Compiled from Fritz et al. (2000) [5] and Fritz et al. (2013) [8].

ELIGIBILITY FOR REPORTING

The principal mission of a population-based cancer registry is to record all cases of cancer that occur in a defined resident population. All malignant tumours as well as non-malignant intracranial and intraspinal CNS tumours (categories III and Xa of ICCC-3-2017) are classified by ICCC, and all classifiable tumours were eligible for inclusion in IICC-3, as specified in the call for data.

Although most cases that are eligible for IICC-3 are recorded in all registries, some groups of neoplasms are excluded or are recorded only sporadically in some registries. The eligibility criteria may be resource-driven, but other considerations, related to disease specification or confidentiality principles, may also play a role. The most common exclusion criteria are linked to the assessed behaviour of tumours.

Non-malignant CNS tumours

Within the target age range of IICC-3, brain tumours constitute a major tumour group. In complete datasets, about one third of these tumours are non-malignant. Non-malignant tumours are those that grow in place but do not spread (benign, code /0), and the term also includes uncertain behaviour (uncertain whether benign or malignant, code /1). Although in situ behaviour, which is also non-malignant, is not coded in ICD-O-3 for the morphologies associated with CNS tumours, the behaviour code /2 was also eligible for inclusion in IICC-3 if reported. Non-malignant CNS tumours can produce similar health effects and require similar therapeutic intervention as malignant tumours. They have an impact on mortality rates, because of their location. Non-malignant CNS tumours were included in the previous IICC volumes.

The reporting of non-malignant CNS tumours by the registries varied, with registries including all of them, selecting only some of them (typically pilocytic astrocytoma), recording them occasionally (non-systematic), or recording none of them. Some registries chose not to submit data on non-malignant CNS tumours to IICC-3, although they collected these data. Table A.9 shows the eligibility for registration of non-malignant CNS tumours, as well as their proportion in each dataset tabulated in IICC-3.

A major contributor to the geographical differences in the incidence rates of CNS tumours may have been the change in the coding of behaviour for pilocytic astrocytoma (ICD-O-3 M-9421) from malignant (code /3) in ICD-O-2 [7] to uncertain (code /1) in ICD-O-3 [5] in 2000. This change resulted in variation in reporting of this tumour. Some registries had stopped collecting information on this tumour if only malignant CNS tumours were eligible for reporting in their registration area (e.g. Australia, South Australia). Others continued collecting this morphology even if no other non-malignant tumours were reportable (e.g. Poland, Kielce). Most registries that continued collecting pilocytic astrocytoma had changed the behaviour code to /1; in others, the behaviour depended on the year of diagnosis: behaviour /3 until implementation of ICD-O-3 and /1 afterwards. The registries in the USA continued to report pilocytic astrocytoma with malignant behaviour, for comparability with older data. Therefore, the number of non-malignant cases, as defined by the behaviour code, may not be comparable across registries or with earlier published data, such as in IICC-2. As can be assessed from Table 10.4, pilocytic astrocytoma may increase the incidence of CNS tumours by up to one third.

The registry-specific tables provide the number of non-malignant CNS cases using the behaviour supplied by the registry, separately for group III and subgroup Xa. In the narratives or in the Editors' comments, further details are provided, such as changes in eligibility criteria during the study period. In Table A.9 the percentage of non-malignant tumours (non-malignant %) relates to the sum of all tumours classified in the diagnostic group III (CNS and miscellaneous intracranial and intraspinal neoplasms) or the subgroup Xa (intracranial and intraspinal germ cell tumours) of ICCC-3-2017. A value of 0.0% of non-malignant tumours should be read in conjunction with information on the collection shown in the adjacent column; it indicates that non-malignant tumours were not eligible, that they were not submitted, or that none were encountered during the study period.

The inclusion of non-malignant CNS tumours results in higher incidence rates. The availability of non-invasive diagnostic techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), or magnetic resonance spectroscopy (MRS), and diagnostic intensity can influence the proportion of non-malignant tumours, so that the incidence rates of non-malignant tumours have a tendency to vary geographically to a greater extent than the incidence rates of malignant tumours.

Myelodysplastic syndromes

Myelodysplastic syndromes (MDS, M-998) were coded to non-malignant (uncertain) behaviour in ICD-O-2 [7] and to malignant behaviour in ICD-O-3 [5]. This change in the coding practices would have affected the completeness of registration of these neoplasms during the study period, in the opposite direction to the pattern described for pilocytic astrocytoma, although the expected impact on the rates would be small because of the rare occurrence of these diseases in the young population.

Carcinoid of appendix

As with other carcinomas, carcinoid tumour is rare in childhood, although its proportion of all neoplasms may start to rise slightly in teenagers. Most carcinoids have malignant behaviour. Carcinoid of appendix (M-8240/1 C18.1) was assigned uncertain behaviour in ICD-O-2 [7] and ICD-O-3 [5]. However, in ICD-O-3.1 [8] the behaviour was changed to malignant. Even before this official change, carcinoid of appendix was coded as malignant by pathologists in some countries (in agreement with the matrix rule F of ICD-O [5, 7]), which may have inflated their incidence rates, not necessarily reflecting differences in cancer risk. The impact of these changes on the presented data may be minimal because this tumour represented much less than 1% of cases in most registries. Percentages that are greater than 1% of the total dataset are highlighted in Table A.11, as indicated for rare cancers in Fig. 4.1.

Skin carcinomas

The subgroup of skin carcinomas (XIe) includes basal and squamous cell carcinomas of the skin. The reluctance of many registries to record these neoplasms springs from the difficulty of ensuring their complete ascertainment, the disproportionate increase in the number of such

tumours with age, and the relative innocuousness of these neoplasms. Therefore, the differences in incidence rates observed in Table 8.29 do not necessarily reflect varying incidence of these cancers, but rather the eligibility criteria applied. These included no registration, non-systematic registration, registration of all except basal cell carcinomas, registration of all except basal or squamous cell carcinomas, registration only if the skin of genitals was the primary site, or complete registration. Table 11.3 shows the distribution of skin carcinomas by primary site.

COLLECTION OF INDIVIDUAL DATA ITEMS

The submitted datasets varied with respect to the coding systems used and the inclusion or completeness of non-mandatory variables.

Coding of tumours

During the registration periods included in IICC-3, the registries may have used one or more coding systems (Table A.10). Most registries submitted cancer data coded to ICD-O-3 [5], and the conversion was done at IARC for other datasets or some records, as indicated in Table A.10. Entire datasets were converted for the registries for which ICD-O-2 [7] is indicated in the column "Coding system submitted", and between 1 and 44 individual histology codes were converted for the registries for which "ICD-O-2" is enclosed in parentheses in Table A.10. The original coding represents the best available information on a given tumour at the time of incidence. A conversion to a later system may influence the eligibility of a tumour for registration; one example is pilocytic astrocytoma, as described above. Sometimes the original code can be converted into more than one new category, and a review of the original medical records is required to recode the case correctly. For example, melanoma coded as M-8742 in ICD-O-2 could be converted either to M-8742/3 or to M-8746/3 in ICD-O-3. Rhabdomyosarcoma coded to M-8900/3 in ICD-O-2 could be converted into three different codes in ICD-O-3: M-8900/3, M8912/3, or M-8921/3. New terms may be introduced, or the behaviour of the existing entities may change from non-malignant to malignant (myeloproliferative neoplasm in ICD-O-3 or Langerhans cell histiocytosis in ICD-O-3.1 [8]). Such changes may affect the comparability of incidence rates over time and between registries, depending on the timing of the uptake of the new classification scheme.

Table A.10 shows the first year of use of a given classification system, as reported by the registries in the IICC-3 questionnaire. These responses were matched with other information (data or supplementary communication with the registries) and modified accordingly if required. Only the classification systems used during the period contributed to IICC-3 are reported, so that the periods over which they were actually used in the individual registries may have been truncated.

Grade

Grade was requested in the call for data as an optional variable. However, because the analyses of the submitted data showed considerable incompleteness, no results are presented for this variable.

Laterality

Cases of retinoblastoma, nephroblastoma, and gonadal tumours were to be optionally provided with an indication of whether they occurred in one side of these paired organs (unilateral tumours) or in both sides (bilateral tumours). Laterality was not collected in many contributing registries and was often incomplete when collected. During the editorial process, multiple diagnoses in the same patients with different laterality were matched using the patient identification number, tumour sequence number, date of birth, and sex of relevant records to identify cases with corresponding diagnoses indicating that both sides were involved in the same patient. If applicable, they were then converted into a single bilateral case and confirmed by the registry. Tables 13.1 to 13.4 show the distribution of laterality by age for the index malignancies in the registries with less than 5% of the cases with missing information. More details can be consulted in the corresponding tables online, in which the laterality distribution is shown for all eligible registries. Thus, comparison of the book and online versions of the laterality tables also indicates the completeness of collection of this data item.

Internal consistency of a cancer record

During the editorial process, the registries were sent lists of records showing various types of queries related to a combination of tumour with age, site with histology, histology with basis of diagnosis, or various rare entities. The choice of validation rules was based on a previous publication [9], with modifications if needed, and additional routines, as described below.

Unlikely combinations of values

Site and morphology

To check for unlikely combinations of site with morphology (histology), a morphological family was assigned to each histology code as in IARCcrgTools [9]. Morphological families group the morphology codes according to the primary sites in which they are likely or unlikely to occur. Every morphology code is classified into a single morphological family. The new morphology codes, created after the definition of morphological families in IARCcrgTools, were also assigned into these families, as shown in Table A.4. In Table A.11, high proportions of cases with unlikely combinations of site and morphology are highlighted (Fig. 4.1). The proportion of unlikely combinations is less than 1% in almost all datasets and does not exceed 5% except in two registries: Libya, Benghazi (10%) and Kuwait. The proportion of cases with unlikely combinations of site and morphology is extraordinarily high in Kuwait (> 50%), as described in the Editors' comments on this dataset in Chapter 6.

Two specific combinations are tabulated separately in Table A.11: carcinomas or neuroblastoma occurring in the brain, as defined in Table 4.4.

Age and cancer type

The provided or calculated age was compared with site, histology, or ICCC-3-2017 category, and the records of cases too young or too old for a given tumour type were sent to registries for verification. The complete set of

Table 4.4. Definition of unlikely or rare entities in IICC-3

ICD-O-3 [5] / ICD-O-3.1 [8]		ICCC-3-2017 (see Chapter 3)	Entity
Morphology	**Topography**	**Subgroup**	
8981			Embryonal carcinosarcoma
8991			Embryonal sarcoma
9072			Polyembryoma
9732			Multiple myeloma
9823			B-cell chronic lymphocytic leukaemia / small cell lymphocytic lymphoma
905*			Mesothelial neoplasms
9500	C70–C71		Neuroblastoma in brain
8810, 8811, 8823, 8830, 8850, 8890, 9120	C40–C41		Rare bone tumours
9180, 9210, 9220, 9240	C49		Rare soft tissue sarcomas
	C70–C72, C75.1–C75.3	XIf	Carcinomas in brain
824*			Carcinoid

*Applies to the whole range of valid morphology codes that start with the digits shown.
Source: Compiled from Fritz et al. (2000) [5] and Fritz et al. (2013) [8].

age–tumour checks is shown in Table 4.3. The proportion of such unlikely records is shown in Table A.11 and is highlighted in red if it represents more than 5% of cases in an individual dataset. The proportion of unspecified carcinomas recorded in infants (age < 1 year) is shown separately and is highlighted if it exceeds 0.5% of the total number of cases, as indicated in Fig. 4.1.

Basis of diagnosis and morphology

The microscopic examination of tissues is required to assign a specific morphology code for a neoplasm. The specific morphology codes that can be assigned on the basis of clinical diagnosis only are shown in Table 4.2 [9]. The proportion of cases with inconsistency between morphology and basis of diagnosis is shown in Table A.11 and is highlighted if it represents more than 10% of cases in each dataset (Fig. 4.1).

Rare entities

The tumour types that are considered to be sufficiently rare to require double-checking for a potential error in recording or coding are listed in Table 4.4, and the proportions in each dataset are shown in Table A.11. Values are highlighted if they represent more than 1% of cases in an individual dataset (Fig. 4.1).

Other validation rules

Other internal consistency rules, listed in Table 4.5, were also applied and resulted in listing the records for validation or correction by the registry. Some of the identified inconsistencies were resolved before data were included in the analyses; the remaining ones were reported, according to Table 4.5.

Incidence rates

The incidence rates were assessed for plausibility overall and by age, sex, and tumour groups, in conjunction with all other information provided by a registry or derived from data. In Table A.9 the overall incidence rates outside the expected limits are highlighted, separately for ages 0–14 years and 15–19 years, as indicated in Fig. 4.1. The incidence rates should be examined in view of all the characteristics of the registry, the eligibility criteria for registration, the data quality, and potential sources of bias, as documented in the registry narratives and Editors' comments in Chapter 6 and summarized in the cited Annex tables.

Conclusions

As presented in this chapter, data users must be aware of the inherent residual differences between the datasets documented in this publication, because they may explain some of the observed variations. While cancer registries continue to harmonize their registration principles and methods to produce internationally comparative results, this volume offers the most complete, timely, and comparable data on the global incidence of childhood cancer.

Table 4.5. Miscellaneous internal consistency checks used in IICC-3

Variables involved	Issue	Management
Age	Inconsistent with dates of birth and diagnosis	Reported in Table A.8
Basis of diagnosis	Unknown	Reported in Table A.9
Behaviour	Metastatic (/6) or unknown (/9)	Resolved
Dates	Incorrect sequence	Resolved
Dates	Missing or incomplete	Reported in Table A.8
Histology	Not listed in ICD-O-3	Resolved
Histology and behaviour	Not listed in ICD-O-3	Included if confirmed
Identification number	Not unique, or missing	Resolved
Laterality	Unknown	Reported in tables in Chapter 13
Sequence number	Unmatched records for a single patient	Resolved
Sex	Unknown	Reported in Table A.8
Sex and morphology	Morphology linked with opposite sex [9]	Resolved
Sex and site	Male with C51–C58 or female with C60–C63	Resolved
Site	Not listed in ICD-O-3	Resolved
Site and behaviour	In situ behaviour (/2) in sites C70–C72	Included

Source: Compiled from Fritz et al. (2000) [5], Fritz et al. (2013) [8], and Ferlay et al. (2005) [9].

REFERENCES

1. United Nations Department of Economic and Social Affairs Population Division (2015). World population prospects, the 2015 revision (DVD edition). Available from: https://esa.un.org/unpd/wpp/Download/Standard/Population.
2. Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD (2005). Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ. 83(3):171–7. PMID:15798840
3. Bray F, Parkin DM (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 45(5):747–55. https://doi.org/10.1016/j.ejca.2008.11.032 PMID:19117750
4. Parkin DM, Bray F (2009). Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 45(5):756–64. https://doi.org/10.1016/j.ejca.2008.11.033 PMID:19128954
5. 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.
6. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al.; IICC-3 contributors (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. Lancet Oncol. 18(6):719–31. https://doi.org/10.1016/S1470-2045(17)30186-9 PMID:28410997
7. Percy C, Van Holten V, Muir C, editors (1990). International Classification of Diseases for Oncology. 2nd ed. (ICD-O-2). Geneva, Switzerland: World Health Organization.
8. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al., editors (2013). International Classification of Diseases for Oncology. 3rd ed., 1st revision (ICD-O-3.1). Geneva, Switzerland: World Health Organization.
9. 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.
10. Bray F, Colombet M, Mery L, Pineros M, Znaor A, Zanetti R, et al., editors (2017). Cancer Incidence in Five Continents, Vol. XI. IARC Scientific Publication No. 166. Lyon: International Agency for Research on Cancer.