CHAPTER 5.

Quality control at the population-based cancer registry

All cancer registries should be able to give some objective indication of the quality of the data that they have collected. The methods available were described in an early IARC Technical Report (Parkin *et al*., 1994) and updated in a pair of papers in 2009 (Parkin and Bray, 2009; Bray and Parkin, 2009). They describe four dimensions of quality: comparability, validity, timeliness, and completeness.

1. Comparability

Comparability of the statistics generated for populations and over time requires the standardization of practices concerning classification and coding of new cases, and consistency in definitions of incidence, such as rules for the recording and reporting of multiple primary cancers occurring in the same individual. The standard for classification and coding of cancer is ICD-O, published by WHO, which provides the standards for coding topography (location of the tumour in the body), morphology (microscopic appearance of the tumour), behaviour (whether the tumour is malignant, benign, or in situ), and grade (the extent of differentiation of the tumour). In addition to that, ICD-O-3 also provides a standard coding scheme for recording the basis of diagnosis and the IARC rules for coding multiple primary cancers. As carcinogenesis is a process that can sometimes take decades, the definition of incidence date is arbitrary, and therefore it is of particular importance to follow the agreed standards. Rules for the definition of incidence date have been given by the European Network of Cancer Registries (http://www.encr.eu/images/docs/recommendations/recommendations.pdf).

In low- and middle-income settings, some objective obstacles might impede following the international standards. For example, the lack of coverage by pathology laboratories, or difficult access to diagnosis, will reduce the percentage of morphologically verified cases, as well as result in postponement of the incidence date according to the standard European Network of Cancer Registries recommendations, prioritizing the date of the first histological or cytological confirmation of the malignancy as the date of incidence.

2. Validity

Accuracy of recorded data is greatly enhanced by consistency checks carried out at the time of data entry, such as those incorporated

into CanReg (see Annex 1). Most registries will also, formally or informally, check on the accuracy of the work of staff by carrying out some sort of re-abstraction (going back to one or more sources, to check on accuracy of recording) or recoding exercises, and acting to correct any obvious deficiencies.

Most registries will report on three statistics that have a bearing on the accuracy of the recoded data. They are:

- the proportion (or percentage) of cases with missing data
- the percentage of cases with a morphologically verified diagnosis (MV%)
- the percentage of cases for which the only information came from a death certificate (DCO%).

2.1 Proportion (or percentage) of cases with missing data

The proportion of cases with unknown values of different data items, such as age or stage, is also an indicator of data quality. An important element to assess here is the proportion of cases with primary site uncertain (PSU%). In addition to the ICD-O code for unknown primary site (C80.9), this category should also include other ill-defined sites.

Some data items can be very difficult to collect in low- and middle-income settings. This can apply, for example, to personal identification number, which then results in more demanding and less accurate linkage procedures. Many LMICs share the problem of unavailability or low quality of mortality data. This can pose numerous problems for a cancer registry, such as under-registration because of the lack of "death certificate only" (DCO) cases contributing to incidence, and inability to calculate the standard data quality indicators (apart from the percentage of cases with a morphologically verified diagnosis [MV%]). The only insight into completeness of cancer registration in the absence of mortality data can be provided by the independent case ascertainment or capture–recapture methods (described below).

2.2 Percentage of cases with a morphologically verified diagnosis (MV%)

Morphological verification refers to cases for which the diagnosis is based on histology or cytology. This is traditionally considered as a sort of "gold standard", with suspicion falling upon the accuracy of diagnosis by other means (although it is questionable whether exfoliative cytology is always more accurate than MRI or CT scan). A high MV% is taken to mean accuracy of diagnosis, whereas a low MV% casts doubt on the validity of the data.

The editorial checks of CI5 include a formal comparison of the MV% (by sex, for the major cancer sites) with a "standard", based on values observed in the same region 5 years earlier. Annex 2 provides the tables with "standard" values of selected data quality indicators, including MV% by country or region, which are used in the CI5 editorial process. Whereas a MV% significantly lower than the expected value may give rise to concern about a lack of validity, it is generally not the cancer registry that can influence the availability of, or use of, pathology services within its area. Usually, in LMICs, the opposite situation – a relatively high MV% – is cause for concern. Collecting data on cancer cases from pathology departments is much simpler than trawling through clinical services or ill-organized hospital archives. A large proportion of cases diagnosed via the pathology department may well suggest defects in case finding and, hence, incomplete registration. Worse, the incompleteness will be biased, with the database containing a deficit of cancers that are not easy to biopsy (e.g. lung, liver, brain, and pancreatic cancer).

2.3 Percentage of cases for which the only information came from a death certificate (DCO%)

DCO cases are those registered on the basis of information on a death certificate, and for which no other information could be traced. As described earlier, the nature of death certificates in LMICs varies widely, from those issued as part of a civil registration of vital events to those generated in a hospital mortuary. However, almost always the accuracy of the diagnostic information is questionable, since the person writing out the certificate may have had little contact with the patient before death and may be ill-informed about how to record cause of death. Thus, if no other clinical record for persons who apparently died of (or with) cancer can be found, there is a reasonable suspicion that the diagnosis was simply wrong. Nevertheless, registry practice demands that such cases are included, but when they comprise a large proportion of cases, the validity of the data is suspect.

Establishing objective criteria of an acceptable DCO% is difficult – it is sensitive to local circumstances, for example availability of death certificates, success in record linkage to the registry database, quality of cause-of-death statements, and facility to trace back cases.

2.4 Internal consistency

Data checks and edits should be applied to newly submitted records to check for item validity, internal

consistency, and inter-record consistency before they are linked with the central database. Such data checks and edits should also be applied to the registry database after any changes have been made.

3. Timeliness

Rapid reporting is often required from the cancer registries. However, for cancer registries (and their clients), a trade-off must be recognized between data timeliness and the extent to which the data are complete. The timeliness depends on the rapidity with which the registry can collect, process, and report sufficiently complete and accurate data. In some countries, such as in the United Kingdom, electronic data capture has expedited the registration process. Some registry networks, such as SEER and the North American Association of Central Cancer Registries, contract their member registries to report data within 22–24 months after the close of a diagnosis year. Some registries use methods such as a delay model estimating the undercount at the time of reporting, or short-term predictions to provide the estimates for the current year.

4. Completeness

Parkin and Bray (2009) distinguish between

- qualitative (or semiquantitative) methods, which give an indication of the degree of completeness relative to other registries, or over time

and

- quantitative methods, which provide a numerical evaluation of the extent to which all eligible cases have been registered.

4.1 Semiquantitative methods

Among the semiquantitative methods, the possibility that a relatively high MV% may represent incompleteness of data collection has already been noted.

A given case may be identified from different sources (hospitals, laboratories, or death certificates), and a large number of different sources per registered cancer case is generally taken to imply that zero sources (i.e. the case was not found in any of them) might be relatively uncommon. The other widely used indicators are:

- mortality-to-incidence ratio
- stability of incidence over time
- comparison of incidence rates with other (similar) populations.

4.1.1 Mortality-to-incidence ratio

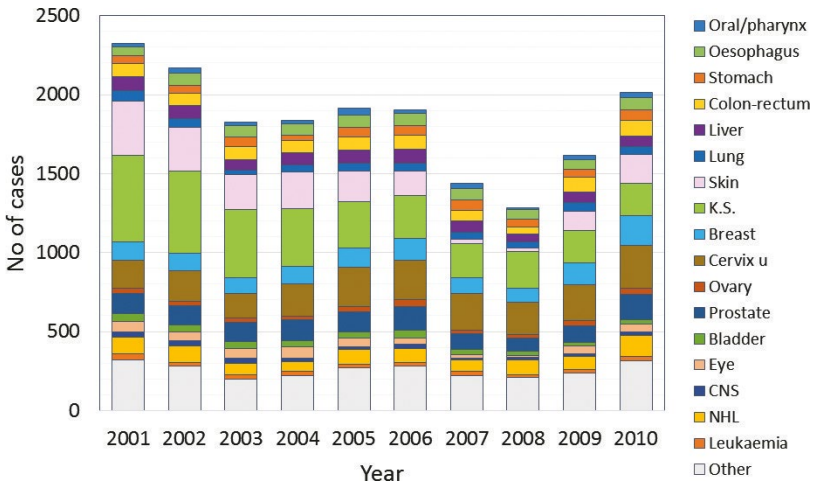
The mortality-to-incidence ratio (M:I) is an important indicator that is widely used – for example, in CI5 – to identify possible incompleteness. It is a comparison of the number of deaths, obtained from a source independent of the registry (usually, the vital statistics system), and the number of new cases of a specific cancer registered in the same time period. Application of this method does require, however, mortality data of good quality (especially with respect to accurate recording of cause of death), so that M:I is approximated by: 1 – survival probability (5 years). This permits objective standards of M:I values to be established, applicable to regions where survival is likely to be more or less similar (see Annex 2). The method cannot be used where there is no comprehensive death registration, or when the cause of death is missing or inaccurate – the situation in almost all countries in Africa, and many of those in Asia.

4.1.2 Stability of incidence over time

It is a simple task for a registry to rapidly check on the number of cases being registered each year. In the absence of marked changes in the population, this can quickly identify potential defects in case finding.

Fig. 5.1 provides an example. There is an obvious deficit of cases for the years 2007, 2008, and 2009, and although this involves most cancer sites, it is especially marked for cancers of the skin.

Fig. 5.1. Number of new cancer cases by site in a cancer registry, 2001–2010. CNS, central nervous system; K.S., Kaposi sarcoma; NHL, non-Hodgkin lymphoma.



CHAPTER 5

4.1.3 Comparison of incidence rates with other (similar) populations

Of course, not all populations will have the same pattern of incidence rates; observing differences is one of the objectives of cancer registration. Nevertheless, it is worth comparing results with those of registries serving a similar population (similar geographically, or of similar ethnic composition) – provided the data from other registries are of good quality themselves – to look for differences. Some variation is to be expected, or may be explicable on the basis of exposure to known risk factors, but a systematic difference (many rates lower than expected) may lead to a suspicion of under-registration.

This method is used by the editors of CI5, where results from each registry are compared with those from a group of registries in the same country (or geographical region) (Annex 2).

4.2 Quantitative methods

Three methods are available to obtain a quantitative evaluation of the degree of completeness of registration:

- independent case ascertainment
- capture–recapture methods
- death certificate methods.

4.2.1 Independent case ascertainment

Comparison of the registry database with sets of cancer cases that have been compiled independently of the cancer registry's case-finding procedures is a particularly useful and objective method of evaluating completeness. It requires record linkage between the cancer registry database and the independent case series, to estimate the numbers of cases in the latter "missed" by the registry. The proportion of eligible patients who are already registered is a direct and quantitative estimate of completeness.

The existence of such files of cancer patients from the registration area – for example, from research studies or surveys – provides an opportunity to evaluate registry completeness that should not be missed.

4.2.2 Capture–recapture methods

Like the numbers of sources per case, this method exploits the fact that cancer registries receive notifications of the same cancer cases from multiple sources. Usually, for this method, sources are grouped into hospital, laboratory (pathology), and death certificate, which are, more or less, independent of each other. The basic idea is that if we know how many cases are notified by one source, a pair of sources, or all three sources, we can estimate how many are notified by none (i.e. were missed). Practically, capture–recapture analysis of completeness requires that record linkage is successfully carried out (so that cases identified by each of the multiple sources are correctly classified). This is no problem for users of CanReg, where the sources of information for each cancer case are brought together. Because of the linked-file structure of CanReg5, this sort of analysis should be particularly straightforward.

4.2.3 Death certificate methods

The death certificate methods depend on the availability of relatively high-quality (complete and accurate) certification of cause of death in the area covered by the cancer registry, and will not be readily applicable in many settings in LMICs. The other two methods can, however, readily be applied.

5. Data quality indices for population-based cancer survival

Unlike incidence data, estimating cancer survival requires a high quality of follow-up information. This is optimally achieved if all-cause mortality data are available as a data source for the registry, and efficient linkage procedures (optimally based on unique identification number) are in place. As in LMICs vital registration systems are often absent, unreliable, or unavailable to the registries, many registries in LMICs have resorted to active follow-up methods. The indices for cancer survival data quality due to exclusion from analysis are frequency of DCO cases and frequency of cases excluded from the study due to lack of any follow-up (Swaminathan *et al.*, 2011). Loss to follow-up is a cause of bias even in registries in high-income countries, as even a small underestimation of deaths can result in overestimation of long-term survival (Brenner and Hakulinen, 2009). In LMICs, with poorly functioning routine health statistics data systems and unavailable mortality data, cancer survival estimates from PBCRs can sometimes provide the only insight into the status of cancer care in the country.

As Skeet noted in *Cancer Registration: Principles and Methods* (Skeet, 1991), "all registries should be able to quote some objective measure of this [ascertainment] rather than relying on received wisdom and pious hope." This is sound advice, which is not always heeded. Reporting of registry results demands some evaluation of their quality, especially as the purpose is almost always to allow a valid comparison of cancer rates and risks, between populations and subgroups and over time, that are not the results of artefacts of the registration process.

Key points

- All PBCRs should be able to provide some objective indication of the quality of the data that they have collected.
- The methods available have been described and updated, and cover four dimensions of quality: comparability, validity, timeliness, and completeness.