

Chapter 6

DATA QUALITY AND INDICES OF RELIABILITY

The cancer registry is a source of information. Quality is a property of the data and a product of the techniques used to create them. The quality of information is a product of the quality of the data and the quality of their presentation. The five main areas of consideration of quality in a PBCR are,

1. **Completeness of coverage:** to obtain information on all cancers diagnosed in the population so as to ensure a high degree of case ascertainment;
2. **Adequacy of data:** to have certain core and critical items of patient information on all cases;
3. **Accuracy of data:** to ensure that data are free from erroneous abstraction, coding, data entry etc.;
4. **Comparability of data:** to adopt information techniques in coding, classification etc., that would assure comparability internally and externally;
5. **Reliability of data:** to ensure reliability of data through periodic audits.

Along with the above objectives, the registries routinely undertake various exercises to ensure that the data they gather and process is of high quality. Based on the publication “Comparability and Quality Control in Cancer Registration” published by the International Agency for Research on Cancer (*Parkin et al, 1994*) and suitably modified so as to be practically applicable for the set-up of the cancer registries under the NCRP, these exercises include:

1. **Details of Sources of Registration:** To minimize duplicate registrations at an early stage and identify number of cases whose residential status including duration of stay remains unknown, a comprehensive account of the various sources of registration is obtained. The latter helps in pinpointing sources and the channels of referral of these resident unknown cases where greater efforts are made to minimize such cases. This exercise also helps in knowing the number of sources of registration per case - an index to assess coverage.
2. **Notification per case:** By notification per case is meant, the different types of reports confirming the diagnosis of cancer in a given case. For example, a patient with a lump in the breast could initially have fine needle aspiration cytology, followed by a biopsy and finally a mastectomy specimen, with all three reports having evidence of malignancy. Notifications per case give an idea of the diagnostic work-up and mechanics of reporting while also providing an indirect measure of completeness of coverage.
3. **Re-abstraction of Random Sample:** All registries regularly re-abstract a ten percent random sample of cases. This is done by supervisory staff, and the core proforma is completed afresh for all these cases, without reference to the original abstracts. Differences between original and re-abstracted data are tabulated and categorized into minor and major disagreements, justifiable disagreements and errors. The above tabulations are prepared by all registries and presented and discussed in the

NCRP workshops every year. Overall such re-abstraction errors have been less than five percent.

4. Range and Consistency Checks: Besides data entry and other checks carried out by the registries, the Coordinating Unit has developed a comprehensive software programme which checks the quality of the data online. This quality control involves range checks, consistency checks, duplicate checks and unknown primary site checks.

Apart from the above steps, within the data itself, there are certain indices of reliability. These include the proportion of microscopic verification, the proportion of cases categorized as “Death Certificate Only” and the Mortality Incidence Ratio.

Proportion of microscopic verification

This is an indicator of the validity of diagnostic information (*Parkin et al, 1994*).

Death Certificate Only

These are cases for which there is no other information other than a death certificate with cancer mentioned as a cause of death. These cases are obtained through the death registration units etc and cannot be matched with any of the registered cases. To some extent the proportion of these cases reflect the proportion of cancer cases that could possibly be missed by the registry. Generally such cases should be minimal.