Pergamon

European Journal of Cancer Vol. 33, No. 6, pp. 812–814, 1997

Printed in Great Britain
0959-8049/97 $17.00 + 0.00

PII: S0959-8049(97)00082-8

Editorial

Reliability of Cancer Registration Data

C.A. Stiller

Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, 57 Woodstock Road, Oxford OX2 6HJ, U.K.

THE PERFECT cancer registry would contain a single complete and accurate record for every case of cancer occurring among residents of the region covered by that registry. The record for each case would be initiated very soon after diagnosis and regularly updated thereafter. No ineligible cases would be included. It is generally recognised, of course, that in practice no registry attains this ideal. But by what margin do registries fall short of perfection and how much does it matter?

The importance of data quality for comparisons between different registries has long been recognised. Over 30 years ago the first volume of *Cancer Incidence in Five Continents* [1] included tabulations showing the percentages of registrations that were histologically verified and that were based on a death certificate only. Since then the chapters and tables on reliability of data in these volumes have increased enormously in size and scope [2], even though one suspects that they have been less consulted than the more enticing tables of incidence rates. However, it is only relatively recently that many registries have carried out studies of the accuracy and completeness of their own data.

Given the continuing high level of public interest in their epidemiology, it is not surprising that several studies of the reliability of registry data have focused on haematological cancers, childhood cancer and especially childhood leukaemia. The latest example is the report on pages 891–894 of this issue by Schouten and colleagues on registration of childhood leukaemia in The Netherlands. These authors compared data collected by the long-established Dutch Childhood Leukaemia Study Group relating to more than 400 cases diagnosed over 4 years with a parallel data set collected by the newly set up Netherlands Cancer Registry. Each of the two registries contained over 95% of the cases that were included in the other. On the crucial, but effectively unevaluable, assumption that notification to the two registries was independent, the completeness of ascertainment for the combined data comfortably exceeded 99%. Records of date of birth, sex, date of incidence and subtype of leukaemia were also compared between the two registries and the numbers of errors were again found to be small.

Probably the most comprehensive study on the accuracy of registry data involved a detailed check on a random sample of 2200 cases recorded by the Scottish Cancer Registration Scheme as having been diagnosed during 1990. The results were published in a monograph [3], a summary paper [4] and a further series of papers relating to individual cancer sites [5–7]. The monograph on the Scottish study [3] together with a literature review carried out at the Bristol Cancer Epidemiology Unit [8] between them summarise, and give references for, nearly all the relevant studies published by 1995. Subsequent studies have included an analysis of data relating to prostate cancer in the Cancer Registry of Norway, in which records were cross-matched with death certificates and hospital diagnostic indices, and a 1% sample was validated from hospital medical records [9]. Deficiency in reporting was estimated as less than 1% and a similarly small proportion of records contained errors which could affect the calculated age and period-specific incidence. The high quality of these data lent weight to the conclusion of a companion paper that, in a country with no organised screening for the disease and a low autopsy rate, there had been an increase in incidence which reflected an underlying increase in risk [10].

Cancer registration data are used for a wide range of purposes. The most basic is provision of estimates of the burden of cancer in the geographical region covered by the registry, or in smaller areas within it. For this, as for all other uses of the data, a high level of ascertainment and effective means of detecting duplicate registrations are required, together with accuracy as regards diagnostic group, date of diagnosis and age. For calculations of subregional incidence, the need for accuracy of address information increases as the size of the area decreases. The most extreme instance relates to investigations of suspected clusters of cancer; then it is essential to validate addresses, in addition to dates and diagnosis, for all cases that appear to form part of the cluster according to any source. As registries accumulate data relating to longer periods, trends in incidence may be studied. Sometimes, the first few years' data are omitted if these are suspected of being incomplete. A uniform level of ascertainment from one period to the

Received 5 Feb. 1997; accepted 5 Feb. 1997.

next is clearly necessary for analyses of trends. This is perhaps particularly important in large data sets, where a fairly small change in ascertainment could give rise to a statistically significant trend in recorded incidence by virtue of the large number of cases studied.

In practice, with good cooperation from data sources, many registries achieve high levels of ascertainment, with figures in excess of 90% being not uncommon. However, difficulties can also occur because of the absence of reliable estimates of the population at risk, over which registries have little control. This can be especially problematic for calculating incidence in small areas or among different ethnic groups, and for analyses of trends [11].

Accurate identifying information is fundamental to the avoidance of duplicate registrations. It is also important for studies involving record linkage; this could be internally, for example, to identify second primary neoplasms, or with other sources, such as death certificates and population registers for calculation of survival. Uniform agreement as to what constitutes a second primary is necessary for studies of second cancers [12]. Under some circumstances, variations in the definition of second primary neoplasms can also substantially affect calculated population incidence rates [13]. International guidelines on the identification and clarification of multiple primaries have recently been produced [14]. While it may be felt necessary to depart from these guidelines for individual analyses, the precise definitions that have been used should be detailed when reporting the results of studies of second primaries.

Several factors have contributed to the growing importance attached to data quality and its formal assessment. Increasingly, registries have a legal or contractual obligation to monitor the reliability of their data. In Norway, for example, where cancer registration has been compulsory since 1952, it has also been obligatory since 1983 for the Cancer Registry to assess the quality of its own data. In the United Kingdom, successive reorganisations of the National Health Service have led to a more explicit contractual framework for cancer registration, and the model core contract for cancer registries includes a requirement for continuing evaluation. Standards are specified for timeliness of registration, completeness of ascertainment, proportion of "death certificate only" cases, proportion of cases with unknown primary site, proportion of histologically verified cases, availability of pathology reports and staging data and coding of diagnosis. The 1995 literature review, under the aegis of the U.K. Association of Cancer Registries, provided the background for future evaluation of data quality [8]. The Association has set up a Quality Assurance Group to monitor the process and coordinate studies, some of which are supported from funds made available for medical audit.

The ever greater use made of cancer registries for both scientific research and health service planning has also contributed to the growth of interest in issues of data reliability. The increase in collaborative studies between registries, both nationally and internationally, has been particularly influential. For the earlier volumes of *Cancer Incidence in Five Continents*, data were submitted in pretabulated form. Nowadays, data are requested in the form of individual records which are checked using a group of standard computer programs. These programs, which check the validity and plausibility of combinations of tumour topography and morphology, sex and age at diagnosis, are available to all registries [15]. In the U.K., at least, they have been accepted as a minimum standard. Not only are uniform standards of data quality essential for collaborative studies, they also make studies from different registries more reliably comparable.

In general, the results of published studies of the quality of cancer registration data have been encouraging. Omissions and other errors have of course been found but, overall, the data, drawn from a wide range of sources, have been shown to be highly reliable and many of the errors have been minor. When relatively low levels of ascertainment have been found, these are likely to reflect delays in registration rather than a simple failure to detect cases [16]. The main value of studies of data quality may well be that, where substantial deficiencies are revealed, they indicate how these may be made good. For example, suppose that two independent registers which store data on the same disease in the same population are cross-matched and each is found to contain only 80% of the cases included in the other; simply combining the two registers, and prospectively their data sources, should increase their ascertainment to 96%. Undoubtedly, as the information in cancer registries is increasingly exploited for both scientific research and health service planning, and the providers of funds to registries take a closer interest in the quality of their data, studies of the reliability of cancer registration will continue to grow in number and importance. This should in turn reinforce the recognition that high-quality cancer registry data depend upon an appropriate level and continuity of funding.

1. Doll R, Payne P, Waterhouse J, eds. *Cancer Incidence in Five Continents*, Vol. I. UICC Technical Report, Berlin, Springer, 1966.
2. Parkin DM, Muir CS. Comparability and quality of data. In Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J, eds. *Cancer Incidence in Five Continents*. Vol VI. IARC Scientific Publications No. 120, Lyon, IARC, 1992, 45–173.
3. Brewster D, Crichton J, Muir C. *Accuracy of 1990 Cancer Registration Data in Scotland*. Edinburgh, Information and Statistics Division, Directorate of Information Services, National Health Service in Scotland, 1994.
4. Brewster D, Crichton J, Muir C. How accurate are Scottish cancer registration data? *Br J Cancer* 1994, **70**, 954–959.
5. Brewster D, Muir C, Crichton J. Registration of non-melanoma skin cancers in Scotland—how accurate are site and morphology codes? *Clin Exp Dermatol* 1995, **20**, 401–405.
6. Brewster D, Muir C, Crichton J. Registration of lung cancer in Scotland. An assessment of data accuracy based on review of medical records. *Cancer Causes Control* 1995, **6**, 303–310.
7. Brewster D, Muir C, Crichton J. Registration of colorectal cancer in Scotland. An assessment of data accuracy based on review of medical records. *Public Health* 1995, **109**, 285–292.
8. Huggett C. *Review of the Quality and Comparability of Data held by Regional Cancer Registries*. Bristol, Cancer Epidemiology Unit, 1995.
9. Harvei S, Tretli S, Langmark F. Quality of protate cancer data in the cancer registry of Norway. *Eur J Cancer* 1996, **32A**, 104–110.
10. Harvei S, Tretli S, Langmark F. Cancer of the prostate in Norway 1957–1991 — a descriptive study. *Eur J Cancer* 1996, **32A**, 111–117.
11. Draper GJ, Kroll ME, Stiller CA. Childhood cancer. *Cancer Surv* 1994, **19/20**, 493–517.
12. Hawkins MM. Multiple primary tumours in population-based cancer registries. *Eur J Cancer* 1996, **32A**, 1279–1280.
13. Filali K, Hedelin G, Schaffer P, *et al.* Multiple primary cancers and estimation of the incidence rates and trends. *Eur J Cancer* 1996, **32A**, 683–690.

14. International Agency for Research on Cancer, International Association of Cancer Registries. *Multiple Primaries*. Lyon, IARC Internal Report No. 94/003, 1994.
15. Parker DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan S. Comparability and quality control in cancer registration. IARC Technical Report No. 19, revised edition. Lyon, International Agency for Research on Cancer, 1996.
16. Alexander F, Ricketts TJ, McKinney PA, Cartwright RA. Cancer registration of leukaemias and lymphomas: results of a comparison with a specialist registry. *Community Med* 1989, **11**, 81–89.

Acknowledgements—I am grateful to Mrs E.M. Roberts for secretarial assistance. The Childhood Cancer Research Group is supported by the Departments of Health for England and Wales and for Scotland.