Acta Oncologica Vol. 33, No. 4, pp. 365–369, 1994

DATA QUALITY AND QUALITY CONTROL OF A POPULATION-BASED CANCER REGISTRY

Experience in Finland

LYLY TEPPO, EERO PUKKALA and MARJA LEHTONEN

Cancer registries should pay great attention to the quality of their data, both in terms of completeness (all cancer patients in the population are registered) and accuracy (data on individual cancer patients must be correct). In addition to technical measures in the data processing, different types of checks and comparisons should be routine practice. Active research policy and ambitious, research-oriented staff with competence in medicine, biostatistics and computer science are essential in terms of maintaining good data quality.

Those who use population-based cancer registry data for planning, research and other purposes expect that the data are of good quality. Completeness of registration means that all patients diagnosed with cancer in the population are included and that all those who are registered are really cancer patients. Accuracy of cancer registry data refers to the correctness of information of individual cancer patients, e.g., place of residence and date of death of the patient, and primary site, date of diagnosis and histological type of the tumour.

Continuous and systematic quality control measures are characteristic of a smooth-running cancer registry (1–3). In this paper the experiences of the Finnish Cancer Registry are described. Even if cancer registries differ in size, in their methods of data collection and in the extent of use of their data, the findings may be generalized to other population-based cancer registries.

Objects and functions of the Finnish Cancer Registry

Finnish Cancer Registry (FCR) was founded in 1952. It is population-based and covers the whole of Finland (population 5.0 mill.). One of its main functions is to collect data on the occurrence of cancer in Finland and to publish cancer statistics. In addition, epidemiological and statistical cancer research is carried out by both the Cancer Registry staff itself and in collaboration with outside researchers.

Data collection, production of cancer statistics. The FCR receives notifications on cancer patients independently from hospitals, pathological and haematological laboratories, physicians, dentists, forensic autopsies and death certificates. Since the late 1980s, an increasing proportion of the notifications are transferred automatically from the computer systems of the laboratories or hospitals to the FCR. Multiple sources of notifications at different phases of the disease (diagnosis, primary treatment, treatment of metastases, autopsy, death certificate) undoubtedly improve the coverage of registration.

At the registry, data are scrutinized visually (e.g., to exclude benign lesions) and then stored in the computer which immediately tells whether the patient is already known to the registry or whether it is a new case. New cases and registered cases with new information (second primary cancer, autopsy report, death certificate, etc.) are transferred to a queue to be coded by trained coders. Coding of new entries usually takes place one to two years from the diagnosis. This activity is supervised by a physician who personally checks all problem cases including those with contradictory information.

Annual cancer statistics are produced on the basis of cases diagnosed during one calendar year. The incidence

Accepted 14 March 1994.
From the Finnish Cancer Registry, Helsinki, Finland.
Correspondence to: Dr Lyly Teppo, Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland.

 ISSN 0284-186X

report for 1991 was published in October 1993. In addition, tabular material is produced in different forms according to requests from epidemiologists, oncologists, administrators, students, journalists, etc.

Use of Cancer Registry data for research. Cancer Registry data are utilized, e.g., to create maps on the occurrence of cancer, cancer trends (for numbers of cases and incidence rates), and social class or occupational group specific risk estimates. This 'descriptive epidemiology' is useful for planning purposes and for formulating (and also testing) hypotheses on risk factors in cancer. The more correct and reliable the incidence figures are, the more meaningful are the observed correlations, differences and trends.

In cohort studies, a group of people with a common characteristic ('exposure') is followed up regarding deaths through the population registry (to calculate the person-years lived by the cohort), and for cancers through the cancer registry. Even though the registration would be somewhat deficient in an unsystematical way, registry data could still be used effectively: the observed and expected numbers of cancers are biased in a similar way, and their ratio (Obs./Exp.) which is commonly used as a measure of cancer risk in a given cohort, is likely to be correct.

In case–control studies, complete coverage is less important. However, the largest possible patient series also for these types of studies can be collected through cancer registries.

If the follow-up for death of the patients in cancer registry files is adequate, population-based survival studies can be made. Accuracy of data on individual cancer patients is critical in these studies.

The material of the FCR has been used several times for various kinds of clinico-pathological studies in which the accuracy of data on individual cancer patients is essential.

The requirements in terms of data quality are thus different for different usages of cancer registry material.

Problems in the completeness of cancer registration

Undiagnosed cancers. Since the development of cancer takes years (if not decades), a large number of cancers will never be diagnosed during the life-time of individuals. Thus, the higher the autopsy rates, the more undiagnosed cancers will be found. A similar effect on the numbers of cancer cases and, consequently, on cancer trends is to be expected when new diagnostic methods are introduced or when mass screenings (such as mammography screening for breast cancer) are organized: ever smaller cancers will be detected. The incidence of cancer means the incidence of diagnosed cancer. Thus, while the increase in the number of cancers due to improvements in diagnostic methodology is technically correct, the increasing incidence trend does not indicate increase in the real occurrence of the cancer in question (4).

Uncertainty of cancer diagnosis. Problems in the coverage of cancer registry data can be conceptual or technical. There are always patients in whom the diagnosis of cancer has not been adequately confirmed. Some of these are never notified to the cancer registry, others may enter the registry files for cancer in a defined organ or at an 'unknown primary site'. This uncertainty in the 'correct incidence' is relevant especially in cancers of the intra-abdominal organs. Old people are more likely to belong to this group than the younger ones. Cancer registries can hardly influence this situation themselves. Along with the improvement in the diagnostic work the proportion of unconfirmed cancer diagnoses and cancers with an unknown primary site decreases and results in erroneous trends for some cancer forms.

The definition of cancer, both in the diagnostic setting (pathologists) and in the clinical work, may change with time (5). A good example is provided by papillomas and carcinomas of the urinary bladder. In the 1950s and 1960s, papilloma was a rather common diagnosis, but subsequently papillomas have been increasingly called papillary carcinomas grade 1, resulting in an erroneous slope of the incidence trend. Another example is the problem of small occult carcinoma-like lesions of the thyroid, breast and prostate. Changes in the interpretation of their malignancy undoubtedly influence cancer statistics, in terms of both completeness and trends.

Underreporting of diagnosed cancers. A varying proportion of diagnosed cancers will never enter the cancer registry files or are reported late, even years after diagnosis. Underreporting may be uneven in terms of geographical region, age of patient and histology of cancer, and it may change over time. Chronic leukaemias, multiple myeloma and related disorders are examples of cancers for which the FCR often receives the first notification by way of the death certificate although the malignant disease had in fact been diagnosed several years earlier. Stable underreporting may remain unnoticed for a long time and thus constitute a serious source of error. Even if the figures and trends look nice and reliable, a deficiency of some 20 to 30% may be hidden behind them.

Cancer registries have developed several ways to overcome underregistration. Stimulation of informants with letters, articles and papers is important. Regular comparison of hospital archives or laboratory files with cancer registry data is a direct way to check the completeness of coverage (6, 7). Monitoring the trends in the numbers of cancer cases or notifications received from different regions, hospitals and laboratories is effective in demonstrating stable reporting or changes in the reporting activity.

Problems in the accuracy of data on individual patients

At the FCR, particular attention is focused on the accuracy of the primary site of tumour and date of diagnosis. These two items usually suffice for proper epidemiolog-

ical studies. If there are problems with coding of these items, additional data are often requested by the FCR from reporting clinicians.

Problems in the definition of primary site may arise, especially in advanced disease or when no operation or endoscopy has been performed. Typical problem areas are cancers in the abdominal cavity. In the FCR, some 2% of all cancers are currently coded to 'primary site unknown'.

Diagnosing cancer may be a long process, and it is not always easy to say what is the date of diagnosis. The rule adopted by the FCR is as follows: the date of diagnosis is the date when the case could be registered even if no further information were to become available. Date of diagnosis has not been a major problem in cancer registration.

Stage is seldom relevant in epidemiology, whereas in clinical and survival studies it is very important. Cancer registries have only limited possibilities to check the correctness of the stage information received. Quite frequently it happens that according to the preoperative evaluation (when the first notification is sent to FCR) the tumour is localized (T1-2N0M0) even though metastases are subsequently found at the pathological examination. Due to this type of inconsistency in data obtained, it is important to obtain reports from the pathologists independently. The postoperative ('real') stage is coded in the FCR.

Detailed histology or cell type of the tumour is open to subjective judgement and interobserver variation, and the nomenclatures used by different pathologists may vary (8). Moreover, tumours are often heterogeneous, and one single biopsy can be misleading. Thus, use of very detailed coding systems is perhaps not justified in cancer registries. Several reclassification exercises have shown that the usefulness of the original histological diagnoses is only limited. These reclassifications have also shown that some of the patients registered as cancer cases had in fact no cancer. Morphological diagnoses are not checked systematically at the FCR. Histological type is being increasingly utilized in epidemiological studies because cancers in one organ constitute a heterogeneous group not only in terms of their clinical behaviour, but probably also of their etiology (9).

Treatment is another item in which cancer registry data cannot be very accurate: extent of operation (curative or palliative), radiation dose or exact and changing regimens of chemotherapy. It was found that if, according to the FCR, certain treatment was coded to be given to patients with cancer of the cervix uteri, this was also the case in the check-up for almost all of the cases, but some patients with a code 'no treatment' had in fact received that treatment.

At the FCR, follow-up for death (or emigration) is a routine procedure through the population registry. No follow-up data for metastases or recurrences are collected. In order to make appropriate survival studies it is important to have a complete follow-up for death (from any cause) of cancer patients. The official cause of death is helpful when calculating cause-specific (corrected) survival rates although subject to several problems. These problems can be avoided when relative survival rates are used.

Technical quality control procedures

There are several technical measures to be used in a routine way which are likely to improve the quality of cancer registry data. The computers can be programmed to detect (and not to accept) invalid codes (e.g., primary site or histology codes not in use), inconsistent combinations of codes and illogical time sequences (dates of diagnosis, commencement of therapy and death). Examples of inconsistent combinations of codes are testis cancer in a female, distant metastases associated with carcinoma in situ type primary lesion, curative surgery in widespread cancer, definitive histology when diagnosis is based on endoscopy or clinical methods only, and code for treatment when diagnosis is made at autopsy only.

Duplicate registration of one individual patient can be effectively avoided at the FCR by using the unique personal registration numbers (given to all residents of Finland).

It is useful to have a system in which unusual combinations of codes can be picked-up and checked. For example, squamous-cell carcinoma of the bone or phaeochromocytoma of the bladder are likely to be incorrect—but not always! Undoubtedly, knowledge of the biological variability of cancer as a disease is important in cancer registration.

Comparison between FCR and hospital discharge registry

Record linkage was made on the basis of the unique personal numbers, between the files of the FCR and the hospital discharge registry (HDR). The HDR covers all patients who have been hospitalized in Finland, and includes the dates of hospitalization and the main diagnoses (ICD codes). Special interest was paid to patients who had been hospitalized with a diagnosis of cancer but who were not found in the FCR. There were 6 034 such patients during the 4-year period 1985–1988. Requests for further information on these patients were sent to hospitals.

In about two-thirds of the cases, the cancer code at the HDR proved to be erroneous, or the diagnosis of cancer appeared to be very unlikely. During the checking process (in 1991–1993), several hundreds of 'missing' cases were notified spontaneously to the cancer registry as late entries (e.g., on the basis of death certificates).

After detailed checking, 1 202 patients remained, about 300 per year, who appeared as if they should have been registered. Of these, 965 were patients with tumours to be included in cancer statistics, whereas 237 were patients with lesions that are not included in statistics (carcinoma in situ of the cervix uteri, papilloma of the urinary bladder,

Table

Total number of cancers diagnosed in 1985–1988 in Finland and the number of cases additionally registered solely on the basis of the information from the hospital discharge registry (HDR), by primary site

Site of cancer	Total No. of cases	Cases based on HDR only	
		n	%
Solid tumours	63 722	579	0.9
Lip	656	9	1.4
Stomach	4 722	7	0.1
Colon	3 811	27	0.7
Pancreas	2 726	7	0.3
Larynx	563	4	0.7
Lung	9 046	29	0.3
Breast	9 296	21	0.2
Corpus uteri	1 965	4	0.2
Prostate	5 053	44	0.9
Testis	257	6	0.3
Kidney	2 486	19	0.8
Urinary bladder	2 495	8	0.3
Skin melanoma	1 871	38	1.9
Other skin[1]	1 921	15	0.8
Eye	184	7	3.8
CNS[2], malignant	1 288	18	1.4
CNS[2], benign	1 339	260	19.4
Thyroid	1 055	4	0.4
Other endocrine gland	137	9	6.6
Unknown site	1 478	14	0.9
Other sites	11 373	29	0.3
Haematological cancers	4 906	386	7.9
NHL[3]	1 561	19	1.2
Hodgkin's disease	483	4	0.8
Multiple myeloma	1 058	95	8.9
Acute leukaemia	863	36	4.2
Other leukaemia	941	232	24.7
Total	68 628	965	1.4
Other disorders	15 765	237	1.5
CIS[4], cervix uteri	892	15	1.7
Papilloma, bladder	108	32	29.6
BCC[5] of the skin	14 296	59	0.4
Polycythaemia vera	231	25	10.8
Myelofibrosis	238	106	44.5

[1] Non-melanoma, non-basal cell carcinoma
[2] Central nervous system
[3] Non-Hodgkin's lymphoma
[4] Carcinoma in situ
[5] Basal cell carcinoma

polycythaemia vera, myelofibrosis and basal cell carcinoma of the skin) (Table).

Among diseases included in cancer statistics there were three groups with a roughly 10% underregistration: benign neoplasms of the central nervous system (mainly among elderly people), chronic lymphatic leukaemia and multiple myeloma (Table). For solid tumours, the coverage was generally rather good. In addition, checks made at pathological laboratories showed that one-third of those 'missing' solid tumours that were reported by the hospitals to be cancers were in fact benign.

More marked underregistration had taken place in lesions not included in cancer statistics. The deficit was about 30% for papilloma of the bladder, 11% for polycythaemia vera and 44% for myelofibrosis (Table).

It appears that the coverage of the FCR for solid cancers is adequate, but great emphasis must be placed on improving the completeness of registration for benign neoplasms of the central nervous system and various haematological disorders.

Concluding remarks

It is difficult to maintain a good cancer registry without an active research programme. If research and researchers are essential elements of a cancer registry, the data quality aspect becomes a natural part of the daily routines of the registry. An enterprising, research-minded staff aiming to be as good as possible under the prevailing circumstances is the best guarantee of effective quality control measures. Quality of cancer registry data depends partly on the competence and experience of its staff which should include expertise in medicine (knowledge on cancer), computer science and biostatistics.

Good relations with practising physicians, health authorities and scientists are important, and all types of conflicts should be avoided. Service is one of the functions of the cancer registry. Production of data requested by researchers and consultation in matters related to the use of cancer registry data and cancer epidemiology generally link the cancer registry and its staff to the scientific community.

An active policy in checking data and asking for further information and more accurate data would be helpful for the data quality. Various types of comparisons and checks against external data sources, both systematically and on an ad hoc basis, give a good impression of the coverage and accuracy of registration. Monitoring of regional incidence trends is an easy way to detect unexpected changes, the reasons for which remain to be elucidated.

The coverage of cancer registration usually refers to whether all patients with a diagnosis of cancer are included. But one should also pay attention to the other aspect: are all patients in the cancer registry files really cancer patients? When individual-based cancer registry material is used for research purposes, for example for clinico-pathological analyses, it may happen that some cancer diagnoses appear to be incorrect or at least very unlikely, or the histological diagnoses should be changed. Thus, feedback from users of the data is invaluable. The cancer registry must be prepared to exclude erroneous entries from its files, just as it should also register new cancer cases even several years after the diagnosis. This also means that cancer statistics can never be considered as 'final' but subject to minor changes even after decades.

In conclusion, the better the quality of data of a cancer registry, the better the possibilities for effective use of these data in planning and research. Conversely, the more active and research-oriented the registry, the better the possibilities of maintaining good coverage and accuracy.

REFERENCES

1. Saxén EA, Teppo L, Hakulinen T. Quality control of cancer registry data. In: Nieburg HE, ed. Prevention and detection of cancer. Part I. Prevention. Vol. 2. Etiology, prevention methods. New York: Marcel Dekker, 1978: 2151–64.
2. Saxén EA. Cancer registry: Aims, functions and quality control. Arch Geschwulstforsch 1980; 50: 588–97.
3. Matsson B, Wallgren A. Completeness of the Swedish Cancer Register. Acta Radiol Oncol 1984; 23: 305–15.
4. Saxén EA. Trends: Facts or fallacy. In: Magnus K, ed. Trends in cancer incidence. Causes and practical implications. New York: Hemisphere Publishing Corporation, 1982: 5–16.
5. Saxén E. Histological classification and its implications in the utility of registry data in epidemiological studies. In: Grundmann E, Pedersen E, eds. Cancer registry. Recent results in cancer research. Vol. 50. Berlin: Springer-Verlag, 1975: 38–46.
6. Kyllönen LEJ, Teppo L, Lehtonen M. Completeness and accuracy of registration of colorectal cancer in Finland. Ann Chir Gynaecol 1987; 76: 185–90.
7. Teppo L. Testicular cancer in Finland. Acta Pathol Microbiol Scand Sect A 1973; (Suppl. 238).
8. Hakama M, Franssila K, Saxén E. Reliability of histopathologic diagnosis of malignant lymphoma. Ann Clin Res 1973; 5: 104–8.
9. Saxén E. Histopathology in cancer epidemiology. The Maude Abbott Lecture. Pathol Annu 1979; 14: 203–17.