PERGAMON

European Journal of Cancer 38 (2002) 414–417

European Journal of Cancer

www.ejconline.com

Reliability of cancer registration data in Scotland, 1997

D.H. Brewster[a,*], D. Stockton[b], J. Harvey[a], M. Mackay[c]

[a]*Scottish Cancer Registry, Information and Statistics Division, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ, Scotland, UK*
[b]*Scottish Cancer Intelligence Unit, Information and Statistics Division, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ, Scotland, UK*
[c]*Scottish Morbidity Records and Standards Unit, Information and Statistics Division, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ, Scotland, UK*

Received 5 July 2001; received in revised form 14 August 2001; accepted 17 October 2001

Abstract

The aim of this study was to assess the reliability of cancer registration data in Scotland following a major re-organisation of the Scottish Cancer Registry. From a random sample of 3500 primary cancers (excluding non-melanoma skin cancers and death certificate only (DCO) registrations) diagnosed between April and September 1997, 3175 (90.7%) had medical records available for scrutiny. Data were re-abstracted by a team of trained medical coders and compared with information registered originally. Reliability was generally high for demographic, diagnostic, and fact of treatment details, but less reliable for grade of differentiation, staging variables and dates of treatment. Some discrepancies probably arose because of differing availability of information at the time of registration compared with the time of re-abstraction. Although data quality is high overall, further improvements might be achieved through continued training, structured recording of information by clinicians in medical records, and continued exploitation of the data. © 2002 Published by Elsevier Science Ltd.

Keywords: Cancer registration; Data quality; Neoplasms; Registries; Scotland

1. Introduction

Population-based cancer registration is an essential part of any rational cancer control strategy [1]. The data provided by cancer registries serve a wide range of functions including public health surveillance, health service planning, evaluation of the impact of interventions on cancer incidence and survival, clinical audit, epidemiological and health services research, and provision of information for many purposes including health promotion and genetic counselling [2]. However, to fulfil all of these functions adequately, the data must be (and must be perceived to be) of high quality [3].

Cancer registration in Scotland was re-organised extensively between the years 1995 and 1997. Formerly, registration was performed by five autonomous regional registries, responsible for collecting a limited data-set, and largely reliant on the manual processing of submitted paper records. Periodically, the regional registries submitted data electronically to a national cancer registry for collation and validation. Following re-organisation, the national cancer registry assumed responsibility for all aspects of cancer registration and for extending the minimum data-set to include information on treatment and (for selected sites) stage of disease for all cases diagnosed from 1 January 1997 onwards. The mode of operation of the registry has been described in detail elsewhere [2]. Briefly, cases are now ascertained from multiple sources, electronically whenever possible, and records are linked by computerised probability matching to identify potential new registrations. Eligibility for registration is then verified by outposted staff referring to medical records; items of data which cannot currently be captured electronically are also abstracted at this time. In a previous study relating to the year of incidence, 1990, we demonstrated that cancer registration data in Scotland appeared to be of high quality [4]. The purpose of this study was to re-assess the quality of the data following what has been a major re-organisation and, in particular, to assess the reliability of new variables in the data-set.

2. Patients and methods

A random sample of 3500 registrations, stratified by health board of residence, was generated from the

* Corresponding author. Tel.: +44-131-551-8903; fax: +44-131-551-8987.
E-mail address: david.brewster@isd.csa.scot.nhs.uk (D.H. Brewster).

0959-8049/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0959-8049(01)00385-9

Scottish Cancer Registry (SCR) by computer. The sampling frame consisted of all primary malignant neoplasms (ICD-10 C00–C96), excluding non-melanoma skin cancers (C44), secondary malignant neoplasms (C77–C80), and death certificate only (DCO) registrations diagnosed between April and September 1997. Permission to access relevant medical records was sought from Medical Directors of hospitals and data were re-abstracted by the Data Quality Assurance (DQA) team of the Information and Statistics Division of the National Health Service (NHS) in Scotland. Re-abstracted information was compared with the information registered originally and discrepancies were identified. For pragmatic reasons, in most cases, only one attempt was made to access medical records; inevitably, some records were unavailable since, for example, they were in use by clinicians at the time of the hospital visit. Records were not sought for patients who had no evidence of attendance at an NHS hospital in their registration record.

3. Results

3.1. Availability of medical records and representativeness of the sample

Of the sample of 3500 registrations, 3175 (90.7%) had medical records available for scrutiny. There were no statistically significant differences between the 3175 registrations assessed and all equivalent 1997 registrations in terms of age ($P=0.144$), sex ($P=0.75$) and major site ($P=0.76$) distribution.

3.2. Identifying/demographic data

There were 10 (0.3%) discrepancies in surname spelling, nine (0.3%) in forename, 41 (1.3%) in date of birth, 137 (4.3%) in address at diagnosis, 182 (5.7%) in postcode of residence at diagnosis, and 50 (1.6%) in general practitioner (GP) practice code. There were no discrepancies in maiden name or sex. Nine of the surname discrepancies were trivial spelling differences which would generate the same Soundex code for record linkage purposes. Of the 182 postcode discrepancies, only 83 (2.6%) would affect small area analyses at the postcode sector level.

3.3. Diagnostic data

Discrepancies in the diagnostic data items are shown in Table 1. Although there was exact agreement on the allocation of incidence date in only 2384 (75%) cases, the re-abstracted incidence date was within 6 weeks of the registered date in 3013 (95%) cases.

There was agreement in coding the major tumour site category (to three digits of ICD-10) in 3078 (97%) cases. The three-digit ICD-10 code discrepancies included 12 non-Hodgkin's lymphomas coded to different codes within the range C82–C85, 18 bladder cancers (C67) re-coded to neoplasms of uncertain behaviour (D41.4), eight stomach (C16) and oesophageal (C15) cancers interchanged, and seven colorectal cancer codes (C18–C20) interchanged. Of those discrepant only at the fourth digit, 264 were coded as unspecified sub-site on the SCR, but were given a specific sub-site on the DQA database. The remainder were general discrepancies between the specific sub-site recorded.

There were 593 (18.7%) discrepancies in the ICD-O(2) morphology code. The majority of these were of little importance. For example, 227 cases recorded as carcinoma, not otherwise specified (M-8010/3) on the SCR were re-coded as neoplasm, not otherwise specified (M-8000/3) on the DQA database. 62 cases registered as adenocarcinoma, not otherwise specified (M-8140/3) were allocated more specific adenocarcinoma codes by the DQA team.

Thirty registrations were recorded as not microscopically verified by SCR and microscopically verified on the DQA database. Conversely, 34 registrations were recorded as microscopically verified on SCR and not microscopically verified on the DQA database. Many (445) of the discrepancies in recording of laterality resulted from recoding to 'not applicable' by the DQA team.

3.4. Grade, stage and oestrogen receptor status

Discrepancies in grade of differentiation, staging variables and oestrogen receptor (ER) status are shown in Table 2. In many instances, discrepancies arose because variables were recorded as 'not known' on one database, but allocated to a specified category or value on the other.

3.5. Treatment data

Discrepancies in treatment data are shown in Table 3. While the fact of treatment was recorded with reasonable reliability for all the treatment modalities, there was less agreement about the dates of commencing treatment.

4. Discussion

In a previous study [4], we showed that the quality of Scottish cancer registration data compared favourably to that reported by the few other cancer registries undertaking such analyses. Overall, the results of this study suggest that the quality of cancer registration data

in Scotland remains high. At 90.7%, the availability of medical records was similar to our previous study, and compares favourably to studies from elsewhere in the UK [5–7]. We believe that our study population was representative of all registrations of malignant neoplasms (excluding non-melanoma skin cancers and secondary malignant neoplasms) incident in Scotland during 1997.

Comparison of the results for demographic and diagnostic variables with those from our previous study [4] suggest that, if anything, data quality has improved slightly since the re-organisation of cancer registration (Table 4). Obviously, we have not been able to compare the reliability of variables which were not collected prior to 1997. It should be noted that our methods differed slightly between the studies. In this study, we excluded non-melanoma skin cancers, secondary malignant neoplasms, and DCO registrations from the sampling frame, and data abstraction was performed by a team of trained medical coders rather than a single medically qualified observer. Nevertheless, we do not believe that these methodological differences invalidate the comparison between our studies. The DCO percentage for all malignant neoplasms, excluding non-melanoma skin cancer decreased from 3.2% in 1990 to 0.4% in 1997, making it less important to pursue such cases in the present study.

The reliability of grade and staging variables is more disappointing, perhaps reflecting that collection of such data was a new departure for the cancer registry in Scotland. Having said that, we are aware from other studies in Scotland that information on grade and staging is often poorly documented in medical records [8], and studies of other cancer registries have demonstrated the difficulty of collecting such data accurately [6,9–11]. The situation may be improved by collection of data prospectively on structured proformas in the context of clinical audit, although the accurate recording of such information is also, of course, essential for clinical purposes.

Although the fact of treatment (within 6 months of diagnosis) was recorded with reasonable reliability in comparison to some other studies [6,7], the date of first treatment for each modality was much less reliable. This may reflect the fact that, in all these kinds of studies,

Table 1
Discrepancies in diagnostic data items

Field	Number of discrepancies	Percentage of discrepancies
Incidence date	772	24.3
Incidence date within 6 weeks	162	5.1
ICD-10 (4 digits)	491	15.5
ICD-10 (3 digits)	97	3.1
ICD-O(2) morphology code	593	18.7
Microscopic verification status	64	2.0
Most valid basis of diagnosis	194	6.1
Method of first detection	109	3.4
Laterality of tumour	537	17.0

Table 2
Discrepancies in grade, stage and ER status

Field	Number of cases analysed	Number of discrepancies	Percentage of discrepancies
Grade classification	3175	629	20.0
Grade of differentiation	3175	293	9.2
Breast clinical T stage	461	84	18.2
Breast clinical N stage	461	79	17.1
Breast clinical M stage	461	47	10.2
Breast pathological tumour size	461	81	17.6
Breast regional nodes examined?	461	51	11.1
Number of nodes examined	461	16	3.5
Any positive nodes?	461	7	1.5
Number of positive nodes	461	11	2.4
Breast ER status	461	22	4.8
Colorectal Dukes' stage	457	61	13.3
Cervix FIGO stage	56	6	10.7

ER, oestrogen receptor; FIGO, International Federation of Gynecology and Obstetrics.

Table 3
Discrepancies in treatment data items

Field	Number of cases analysed	Number of discrepancies	Percentage of discrepancies
Surgery	All	125	4.0
Date of first surgery	1769	292	16.5
Treated with radiotherapy	All	60	1.9
Date of first radiotherapy	954	231	24.2
Chemotherapy	All	55	1.7
Date of first chemotherapy	689	106	15.4
Hormone therapy	All	52	1.6
Date of first hormone therapy	559	98	17.5

Table 4
Comparison between quality of 1990 and 1997 cancer registration data

Field	Discrepancy percentage	
	1990	1997
Surname	0.9	0.3
Forename	1.0	0.3
Sex	0.4	0.0
Date of birth	1.3	1.3
Postcode sector of residence	3.3	2.6
Incidence date within 6 weeks	11.0	5.1
Microscopic verification	7.7	2.0
ICD-9/10 code (3 digits)	5.4	3.1
ICD-O(1/2) morphology code	28.3	18.7

some discrepancies arise simply because different information is available to the person making the original registration and the person re-abstracting data for comparison.

In conclusion, we have been reassured that the reorganisation of cancer registration in Scotland has not had an adverse impact on data quality, and if anything, data quality has been enhanced. We hope that the results of this and previous studies by our own and other cancer registries will help to dispel the myth that it is not possible to achieve cancer registration data of adequate quality to fulfil many important functions in the arena of cancer control [12]. There will always, however, be scope for improvement and our study underlines the need for continuing investment in training of registry staff, as well as ready access to high quality, preferably structured medical records. The collection of high quality data is not an end in itself and, as others have observed, exploitation of cancer registry data is also a vital component of any quality assurance programme [13].

Acknowledgements

We are grateful to Medical Directors for granting permission to access medical records for this study, and to staff of the Medical Records Departments, and our own Cancer Registration Officer colleagues for facilitating the study. The data for the study were collected by Val Borland, Gillian Boyle, Laura Dobbie, Vi Duguid, Kjersti Fergusson, Sophie Houston, Chris Jones, Fiona MacKenzie, Etta Shanks, Tim Varley and Mary Virtue of the Information & Statistics Division Data Quality Assurance team. Finally, we are grateful to the anonymous reviewers for their constructive comments on the manuscript.

References

1. Muir CS, Démaret E, Boyle P. The Cancer Registry in cancer control: an overview. In Parkin DM, Wagner G, Muir CS, eds. *The Role of the Registry in Cancer Control. IARC Scientific Publications No. 66*. Lyon, International Agency for Research on Cancer, 1985, 13–26.
2. Scottish Cancer Intelligence Unit. *Annual Report 1998/99*. Edinburgh, ISD Scotland Publications, 1999 (http://www.show.scot.nhs.uk/isd/scottish_health_statistics/subject/SCIU/1999.pdf).
3. Skeet RG. Quality and quality control. In Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, eds. *Cancer Registration Principles and Methods. IARC Scientific Publications No. 95*. Lyon, International Agency for Research on Cancer, 1991, 101–107.
4. Brewster D, Crichton J, Muir CS. How accurate are Scottish cancer registration data? *Br J Cancer* 1994, **70**, 954–959.
5. West RR. Accuracy of cancer registration. *Br J Prev Soc Med* 1976, **30**, 187–192.
6. Gulliford MC, Bell J, Bourne HM, Petruckevitch A. The reliability of Cancer Registry records. *Br J Cancer* 1993, **67**, 819–821.
7. Pollock AM, Vickers N. Reliability of data of the Thames cancer registry on 673 cases of colorectal cancer: effect of the registration process. *Quality in Health Care* 1995, **4**, 184–189.
8. Howard GCW, Thomson CS, Stroner PL, Goodman CM, Windsor PM, Brewster DH. Patterns of referral, management and survival of patients diagnosed with prostate cancer in Scotland during 1988 and 1993: results of a national, retrospective population-based audit. *Br J Urol* 2001, **87**, 339–347.
9. Polissar L, Feigl P, Lane WW, Glaefke G, Dahlberg S. Accuracy of basic cancer patient data: results from an extensive recoding survey. *J Nat Cancer Inst* 1984, **72**, 1007–1014.
10. Liu WL, Kasl S, Flannery JT, Lindo A, Dubrow R. The accuracy of prostate cancer staging in a population-based tumor registry and its impact on the black-white stage difference (Connecticut, United States). *Cancer Causes Control* 1995, **6**, 425–430.
11. Schouten LJ, Langendijk JA, Jager JJ, van den Brandt PA. Validity of the stage of lung cancer in records of the Maastricht cancer registry, The Netherlands. *Lung Cancer* 1997, **17**, 115–122.
12. Bell CMJ, Lawrence G, Pheby DFH, Smith J, Coleman MP. The role of cancer registries. *Clin Oncol* 1995, **7**, 143–144.
13. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 1994, **33**, 365–369.