

nine isoniazid resistant isolates. DNA sequencing showed that eight of the nine isolates had a point mutation, deletion or insertion of one to three bases only.

Our probe for the *M tuberculosis* katG gene hybridised with all 26 smear and culture positive clinical isolates of *M tuberculosis*, indicating that the katG gene was not deleted completely in any of the 20 isoniazid resistant strains. These three studies taken together suggest that complete deletion of the katG gene does not seem to be the underlying mechanism of isoniazid resistance. Our probe, though specific for *M tuberculosis*, cannot be used as a marker for isoniazid susceptibility.

The authors wish to acknowledge the support and advice provided by Professor P N Tandon, Professor of Neurosurgery, AIIMS; Dr Rama Mukherjee, Deputy Director, NII, Dr V M Katoch, Deputy Director, JALMA, and Dr Bimal K Das, Department of Microbiology, AIIMS, during the course of this work. Financial support from the Department of Biotechnology is gratefully acknowledged.

1 Cole ST. Mycobacterium tuberculosis: drug resistance mechanisms. *Trends Microbiol* 1994;2:411-20.

- 2 Middlebrook G. Isoniazid resistance and catalase activity of tubercle bacilli. *American Review of Tuberculosis* 1954;69:471-2.
- 3 Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase-peroxidase gene and isoniazid resistance of *M. tuberculosis*. *Nature* 1992;358:591-3.
- 4 Zhang Y. Genetic basis of isoniazid resistance of *M. tuberculosis*. Proceedings of the 9th Forum in Microbiology. Amsterdam: Elsevier Science 1994:143-9.
- 5 Kent PT, Kubica GP. *Public health mycobacteriology. A guide for the level III laboratory*. Atlanta: Centers for Disease Control, 1985.
- 6 Baron EJ, Peterson LR, Finegold SM. Mycobacteria. In: Bailey Scott's *Diagnostic microbiology*. 9th edn. St Louis: Mosby, 1994:590-633.
- 7 Jaber M, Rattan A, Verma A, Tyagi J, Kumar R. A simple method of DNA extraction from Mycobacterium tuberculosis. *Tuber Lung Dis* 1995;76:578-81.
- 8 Heym B, Zhang Y, Poulet S, Young D, Cole ST. Characterization of katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of *M. tuberculosis*. *J Bacteriol* 1993;175:4255-9.
- 9 Zhang Y, Garbe T, Young D. Transformation with katG restored isoniazid sensitivity in *M. tuberculosis* isolates resistant to a range of drug concentrations. *Mol Microbiol* 1993;8:521-4.
- 10 Stoeckle MY, Guan L, Riegler N, Weitzman I, Kreiswirth B, Kornblum J, et al. Catalase-peroxidase gene sequences in INH susceptible and resistant strains of *M. tuberculosis* from New York city. *J Infect Dis* 1993;16:1063-5.
- 11 Altamirano M, Marostenmaki J, Wong A, FitzGerald M, Black WA, Smith JA. Mutations in the catalase-peroxidase gene from INH resistant *M. tuberculosis* isolates. *J Infect Dis* 1994;19:1162-5.

*J Clin Pathol* 1996;49:947-949

## Benefits and limitations of pathology databases to cancer registries

D H Brewster, J Crichton, J C Harvey, G Dawson, E R Nairn

### Abstract

In order to assess the benefits and limitations of pathology databases to cancer registries, computerised pathology records of malignant neoplasms diagnosed during 1992 were obtained for a defined area of Scotland for which pathology data were not routinely being used for cancer registration. Apparently 'missed' cancer registrations were identified by computerised probability matching with cancer registration records and their eligibility for registration was determined by reference to medical records, or when these were unavailable, by reference to the text of the original pathology report in conjunction with the local Community Health Index (to establish residency at the time of diagnosis). Misclassifications of site or incidence year were not regarded as 'missed' cases. Of 218 apparently 'missed' cancer registrations identified from computerised pathology records, 133 (5.7% of the revised total number of registrations for the study area in 1992) should have been registered. A further 14 cases were already registered but with misclassified site, morphology and/or behaviour codes. Ascertain-

ment of cases by the Scottish Cancer Registration Scheme seems to be high for most sites. Pathology databases represent a useful additional source of cases but the fact that 71 apparently 'missed' cases were found to be ineligible for registration as independent primary malignant neoplasms suggests that unverified computerised pathology data should not be used uncritically nor independently for cancer registration purposes. (*J Clin Pathol* 1996;49:947-949)

Keywords: neoplasms, pathology, registries.

High quality cancer registration data are an essential prerequisite of any rational cancer control strategy.<sup>1</sup> As about 81% of registrations in Scotland are verified histologically (unpublished data, 1994), pathology data represent a potentially excellent source of case ascertainment and offer the prospect of increasing the validity of diagnoses and the accuracy of information recorded about morphology.<sup>2</sup>

In order to assess the benefits and limitations of computerised pathology data to a cancer registry, we studied a defined area of Scotland served by a single pathology laboratory whose data were not already being used for cancer

Scottish Cancer Registry, Information and Statistics Division, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ  
D H Brewster  
J Crichton  
J C Harvey  
G Dawson

Department of Pathology, Crosshouse Hospital, Kilmarnock, Ayrshire KA2 0BE  
E R Nairn

Correspondence to:  
Dr D H Brewster.

Accepted for publication  
12 August 1996

registration purposes during 1992. In this laboratory, consultant pathologists are responsible for diagnostic coding. The study area has a population of approximately 377 000 and 2202 registrations of malignant neoplasms were recorded for the area in 1992. Registration in the area was primarily based on notification forms supplied by coding clerks from the department of medical records at the time of hospital discharge (either from local hospitals or tertiary referral centres). In addition, potential registrations were continually sought from death records with any mention of cancer supplied by the General Register Office (Scotland). Although the Scottish Cancer Registration Scheme collects information on neoplasms of uncertain behaviour and in situ neoplasms, our study was restricted to *malignant neoplasms only*. Secondary malignant neoplasms are only registered in cases where the primary site is not established. At the moment, due to resource constraints, only one basal cell carcinoma is collected per person.

### Methods

Computerised pathology records relating to the year 1992 were downloaded from the local Pinnacle pathology system on to a dBase file. On the basis of the SNOMED morphology code range 8\*\*\*-9\*\*\*, and fifth digits (behaviour codes) 3 (primary malignant neoplasms), 6 (secondary malignant neoplasms) and 9 (malignant neoplasms, uncertain whether primary or secondary), 5174 records were selected. To facilitate

record linkage with the cancer registration database, ICD-9 site codes for each record were generated manually based on the textual 'part description' and the SNOMED topography code. At this stage, obvious duplicates and secondary tumours clearly associated with a primary site recorded in the same individual were discarded. The remaining 2268 records were linked by computerised probability matching<sup>2</sup> to the cancer registration database (all years) to identify apparently 'missed' registrations. The output of this linkage was checked manually to eliminate secondary tumours clearly associated with previously registered primary tumours and same site duplicate tumours based on the International Agency for Research on Cancer (IARC) multiple primary rules.<sup>4</sup> In a previous report<sup>5</sup> we quantified the extent of misclassification of diagnosis and incidence year in the Scottish cancer registration database. For the purposes of this study, we did not regard misclassifications of site or incidence year as 'missed' cases. Eligibility for registration of the remaining 218 apparently 'missed' cases was determined by reference to medical records, or when relevant medical records were unavailable (20 cases) by reference to the text of the original pathology report in conjunction with the local Community Health Index held by the Primary Care Division of the Health Board (to establish residency at the time of diagnosis). In five cases, with no records on the Community Health Index, residency was established by reference to death certificates. Estimated completeness of

Table 1 Estimated completeness of case ascertainment by the Scottish Cancer Registration Scheme, based on an independent comparison with pathology data, 1992

ICD-9 code	Site	No. of cases (x)	No. of completely 'missed' cases (y)	Estimated % completeness (x/(x+y)×100)	95% Confidence intervals	
					Lower	Upper
140-149	Lip, oral cavity and pharynx	39	1	97.5	96.7	98.3
150	Oesophagus	58	0	100.0	100.0	100.0
151	Stomach	89	1	98.9	98.7	99.1
153, 154	Colon, rectum	192	20	90.6	90.3	90.8
155, 156	Liver, gall bladder	30	0	100.0	100.0	100.0
157	Pancreas	57	1	98.3	97.8	98.7
152, 158, 159	Small intestine, retroperitoneum, other digestive sites	14	0	100.0	100.0	100.0
160	Nasal cavities, middle ear, sinuses	1	1	50.0	1.0	99.0
161	Larynx	21	1	95.5	93.6	97.3
162	Trachea, bronchus, lung	360	2	99.4	99.4	99.5
163	Pleura	10	0	100.0	100.0	100.0
170, 171	Bone, connective tissue	9	0	100.0	100.0	100.0
172	Skin (melanoma)	41	0	100.0	100.0	100.0
173 (M-807)	Skin (squamous cell)	103	12	89.6	89.0	90.1
173 (M-809)	Skin (basal cell)	330	46	87.8	87.6	87.9
173 (M-8832)	Skin (dermatofibrosarcoma)	1	1	50.0	1.0	99.0
174, 175	Breast (female/male)	247	13	95.0	94.8	95.2
179	Uterus	2	1	66.7	35.9	97.5
180	Cervix uteri	33	6	84.6	82.8	86.4
182	Corpus uteri	24	2	92.3	90.3	94.3
183	Ovary	46	1	97.9	97.3	98.5
184	Other, unspecified genital organs	11	0	100.0	100.0	100.0
185	Prostate	86	6	93.5	93.0	94.0
186	Testis	11	0	100.0	100.0	100.0
187	Penis	2	0	100.0	100.0	100.0
188	Bladder	115	2	98.3	98.1	98.5
189	Kidney	29	0	100.0	100.0	100.0
190	Eye	2	0	100.0	100.0	100.0
191, 192	Brain, central nervous system	30	0	100.0	100.0	100.0
193	Thyroid gland	6	1	85.7	75.9	95.5
194	Endocrine glands	1	0	100.0	100.0	100.0
196-199	Secondary sites	83	0	100.0	100.0	100.0
200 + 202	Non-Hodgkin's lymphoma	53	8	86.9	85.8	88.0
201	Hodgkin's disease	8	1	88.9	82.0	95.7
203	Multiple myeloma	16	2	88.9	85.5	92.3
204-208	Leukaemia	42	4	91.3	90.1	92.5
140-208	All malignant sites	2202	133	94.3	94.3	94.3

case ascertainment by the cancer registry was calculated by dividing the existing number of registrations by the sum of the newly identified and existing registrations. The 95% confidence intervals of these estimations were calculated using the normal approximation.

### Results

Of 218 apparently 'missed' cancer registrations identified from computerised pathology records, 133 (5.7% of the revised total number of registrations for the study area in 1992) were eligible for registration. Overall, this implies that approximately 94% of all cases are ascertained by the cancer registry, although there are variations in this estimate of completeness by site (table 1). Estimates of completeness are high for most of the major sites, whereas the lower estimates (for example, 50%) relate to small numbers of cases and therefore have wide confidence intervals.

A further 14 cases were already registered but with misclassified site, morphology and/or behaviour codes.

The remaining 71 apparently 'missed' cases which were not eligible for registration as independent primary malignant neoplasms were discarded for a variety of reasons including one or more of the following: recurrences of or metastases from previously registered neoplasms; failure of the SNOMED topography code to represent the presumed site of origin of the neoplasm; errors in SNOMED coding (for example, solar keratosis coded as squamous cell carcinoma of the skin); subsequent revisions of diagnosis; and residency outside the study area at the time of diagnosis.

### Discussion

The results of our study suggest that ascertainment of cases by the Scottish cancer registration scheme is very efficient for most sites and compares favourably with estimates of completeness published by other cancer registries.<sup>6</sup> It is also important to remember that some cases missed during life are eventually registered as a result of death certification. It is reassuring to find high levels of ascertainment for the major sites of cancer, which in conjunction with a previous assessment of the accuracy of information held by the cancer registry,<sup>5</sup> suggests that national estimates of incidence are likely to be fairly robust. Missed registrations of rarer tumour sites have a more dramatic effect on estimates of incidence and demonstrate more clearly the potential benefits of using pathology data to ascertain cases.

As our independent comparison has been with pathology data from a district general hospital, we have not been in a position to identify cases diagnosed by specialist pathology departments (for example, neuropathology) in tertiary centres, nor cases diagnosed without pathology which have been missed. Thus our estimates of completeness for some sites (for example, cerebral tumours) may be overopti-

mistic. Our application of the IARC multiple primary rules may have also resulted in an overoptimistic estimate of completeness by excluding some genuine independent primaries from the 'missed' cases. However, the occurrence of same site multiple primaries is uncommon for most sites.<sup>7</sup>

It has been observed previously that reliance by cancer registries on pathology data alone may lead to loss of detail about subsite<sup>8</sup> and failure to identify cases of recurrent disease or to determine the correct incidence date.<sup>9</sup> Our experience concurs with these observations and suggests that while computerised pathology data represent an excellent source of potential cancer registrations, details should be verified ideally by reference to medical records (resources permitting), the full text of the pathology report or linkage to other evidence, such as hospital discharge data.

Histopathologists have been described as fairy godmothers to cancer registries<sup>1</sup> and, while their primary concern will always be the management of individual patients, they should not underestimate their potential contribution to maintaining and improving the quality, comparability and utility of cancer registration data.<sup>2,10</sup> In particular, meticulous attention to recording and coding of diagnostic details (reinforced by quality assurance schemes), and notification of diagnostic revisions, would help to eliminate some of the most serious errors in the cancer registration database. Subject to the quality of tissue specimens received, the provision of prognostic indicators (such as tumour stage and grade) to cancer registries will also become increasingly important to enhance the validity of comparative survival analyses.

In summary, we have shown that ascertainment of malignant neoplasms by the Scottish Cancer Registration Scheme seems to be high for most sites. Computerised pathology data represent a useful additional source of cases, but should not be used uncritically nor independently for cancer registration purposes.

- 1 Joslin CAF. National cancer control and cancer registration [editorial]. *Br J Cancer* 1990;62:882.
- 2 Brewster D. Improving the quality of cancer registration data. *J R Soc Med* 1995;88:268-71.
- 3 Kendrick S, Clarke J. The Scottish Record Linkage System. *Health Bulletin (Edinburgh)* 1993;51:72-9.
- 4 Muir CS, Percy C. Classification and coding of neoplasms. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, eds. *Cancer registration: principles and methods*. Lyon: IARC, 1991:64-81.
- 5 Brewster D, Crichton J, Muir CS. How accurate are Scottish cancer registration data? *Br J Cancer* 1994;70:954-9.
- 6 Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL. *Comparability and quality control in cancer registration*. IARC Technical Report No. 19. Lyon: IARC, 1994.
- 7 Boice JD Jr, Curtis RE, Kleinerman RA, Storm HH, Jensen OM, Jensen HS, et al. *Multiple primary cancers in Connecticut and Denmark*. National Cancer Institute Monograph No 68. Bethesda: National Cancer Institute, 1985.
- 8 Lapham R, Waugh NR. An audit of the quality of cancer registration data. *Br J Cancer* 1992;66:552-4.
- 9 Kee F, Gorman D, Odling-Smee W. Confidence intervals and interval cancers...needles and haystacks? *Public Health* 1992;106:29-35.
- 10 Davies TW. Cancer registration and pathology. *Bulletin of the Royal College of Pathologists* 1995;92:17.