Brit. J. prev. soc. Med. (1976), **30**, 187-192

Accuracy of cancer registration

R. R. WEST

Department of Community Medicine, Welsh National School of Medicine

West, R. R. (1976). *British Journal of Preventive and Social Medicine*, **30**, 187-192.
Accuracy of cancer registration. In South Wales cancer registration is done principally by means of the Hospital Activity Analysis. Altogether 1460 hospital records of cancer patients (19% of the 1972 registrations received by May 1973) were studied and the principal items of information required for cancer registrations by the Office of Population Censuses and Surveys were copied and subsequently compared with the corresponding registrations at the Welsh Hospital Board's cancer bureau. Differences between these 're-registrations' and the original registrations were analysed item by item. There were 234 registrations with errors in the diagnostic summary (although 110 of these would cause misclassification only under the fourth digit of the *ICD* code), 164 with errors in date of birth (36 of which would cause classification in the wrong WHO age group) and 198 with errors in the date of registration (112 of which were wrongly ascribed to the year 1972). Error and omission rates were particularly high for NHS number, occupation, place of birth, and histology.

Cancer registration constitutes one of a number of vital statistics, which, as distinct from death certification, is helpful both in planning therapeutic services and in epidemiological research. Studies have reported on the accuracy of hospital morbidity data (Lockwood, 1971) and more specifically on the accuracy in notification of the cause of death (Heasman and Lipworth, 1966; Alderson and Meade, 1967) and on related statistics as for example the accuracy of the occupational description of coal miners (Heasman, Liddell, and Reid, 1958). Although much resource is expended internationally on the collection of cancer statistics and on cancer research, little is known about the accuracy of data collected by cancer registers (Hansluwka, 1975). The study described here is an estimate of the accuracy of the information collected by the South Wales Registry for the Office of Population Censuses and Surveys (OPCS). In South Wales cancer registration is done principally by means of the Hospital Activity Analysis (HAA), the information system in which certain patient identification data are abstracted from hospital notes and collected by regional health authorities (West, 1973a). With this method, introduced in 1972, registration is effected very much sooner after diagnosis and the total registration rate has risen by 20% to over 3·6 per thousand a year (West, 1973a; Welsh Office, 1974).

Method

A one in five sample of 1972 registrations received by May 1973 was selected for this single observer study. The selected cases were identified by the hospital and hospital case record number and were 're-registered' at hospital directly from the hospital notes. The information so obtained was subsequently compared with the corresponding record at the central registry. The principal items of information required by the OPCS for cancer registration were studied and the difference between registrations and study 're-registrations' are reported. It is appreciated that 're-registrations' may not be 'correct' in all cases, but the frequencies of differences between 're-registrations' and the original registrations are considered as indicative of the frequencies of possible errors in each item of information and the magnitude of inaccuracy in cancer registration.

Results

In May 1973 7500 cases of cancer had been registered by the South Wales Cancer Registry for the year 1972. Of these, 1800 registrations (24% of 7500) from 61 hospitals in 10 hospital management committees (HMCs) were randomly selected from the 1972 register and 1460 (19% of 7500), 724 males and 736 females, were successfully traced and 're-registered'. The analysis of differences between 1460 're-registrations' and the corresponding registrations was done for each hospital (and for each HMC) to study interhospital variations and hospital specific errors and omissions. However, in this paper the results of the whole sample are treated together.

Number of Patients' Notes Found

Only 1460 (81%) of hospital notes were found out of the 1800 cancer registrations selected for this study; the proportion found in each HMC ranged between 75 and 100% and the proportion found in each hospital ranged more widely (one large general hospital found only 58%). Such failure to find 19% of hospital notes is indicative of malfunction of hospital records' departments and has been found in many studies. Analysis of errors and omissions in the cancer registration data studied by hospital (and by HMC) showed no significant association with the proportion of notes found in each hospital (or HMC).

Number of Duplicate Registrations

Altogether 27 duplicate registrations were identified (2% of 1460). Of these 18 were duplicates from within the area covered by the registry, but nine were first diagnosed (and registered) in other regions or in other countries. Approximately 2% duplicate registrations are normally found (and duly cancelled) at the first routine annual follow-up by the cancer bureau; this study predated the routine follow-up.

Errors and Omissions in Patients' Names

There were 310 registrations in which the patients' names were incorrectly or incompletely transcribed from the hospital notes to the cancer register (Table I). The 12 surname errors (0·8%) were either spelling mistakes or an incorrect substitution of more common similar names. Most of the 151 (10·3%) errors and omissions of second and subsequent forenames and of 117 errors and omissions of maiden names (18% of married women) were omissions. Analysis by hospital (and by HMC) showed that it was the practice of certain hospitals (and HMCs) to enter only the surname and first forename on cancer registrations even when other forenames were recorded in hospital notes.

Table I

FREQUENCIES OF ERRORS AND OMISSIONS AMONG 1460 CANCER REGISTRATIONS: SOUTH WALES, 1972

Parameter	Number of Errors and Omissions		Percentage of 1460	
Duplicate registrations	27		2	
Name	193		13	
Surname		12		1
First name		30		2
Second or subsequent forename		151		10
Maiden name (of married women)	117		(18% of 662)	
Address (sufficient to lead to wrong area of residence coding)	24		2	
Date of birth	164		11	
Date of registration	198		14	
Diagnosis	234		16	
Occupation	481		32	
(also omissions or only 'housewife', 'child', 'retired', etc. on hospital notes		608		41
Place of birth	276		19	
(also omissions on hospital notes)		459		31
Histology	264			
(also omissions on hospital notes)		272		19
NHS number	170		12	
(also omissions on hospital notes)		1229		84

Errors and Omissions in Address

In only two hospitals were all minor mistakes and 'unimportant' omissions counted, giving very high frequencies: for a hospital in an urban area it was seven in 40 registrations and in a rural area 15 in 42. A less critical threshold was set for the main study, so that only major errors or omissions, likely to result in allocation of cases to wrong areas of residence, were counted. Twenty-four such errors were identified and these included temporary addresses instead of permanent addresses and inclusions of villages and small townships with neighbouring major towns.

Errors in Date of Birth

There were 164 registrations (11·2% of 1460) with dates of birth recorded by the cancer register differing from those in the corresponding hospital notes. There was no significant deviation in the age distribution of dates of birth errors from that of cancer registrations as a whole. There were 20 errors in days of birth, 21 errors in month of birth, and 51 omissions of days and months; where only the age in years was given to the cancer registry yet the full date of burth was found in hospital notes. Of the 72 errors (4·9%) in the year of birth, 36 (2·5%) would have caused tabulation under the

wrong WHO age grouping. There were seven registrations for which dates of registration were entered in place of dates of birth.

ERRORS IN DATE OF REGISTRATION

There were 198 errors in the date of registration (13·6% of 1460 registrations), 112 (7·6%) of which were ascribed wrongly to the year 1972. In Table II those registrations incorrectly dated are listed by the years when they should have been registered for seven broad classifications of diagnosis (using the *International Classification of Diseases* (*ICD*), 8th edition). When the incidences of errors by diagnostic group are compared with the incidences of cancer registrations, it appears that there are relatively few late registrations of previously diagnosed (and treated) cancers of the digestive organs (*ICD* codes 150-159) and the respiratory system (160-163), cancers with poor post-registration survival, but an excess of cancers of the genito-urinary organs (180-189) and of benign and unspecified neoplasms (210-239), cancers with good post-registration survival (Welsh Hospital Board, 1974). A second analysis of errors in dates of registration was done to find where the information relating to the correct date of registration was. For more than half (113) the information relating to the first cancer diagnosis (and treatment) was in hospital notes of the hospitals making the late registrations. There were 59 registrations for which the earlier cancer information was available in hospital notes of another hospital in the South Wales cancer registry region, but most of these (36) were in registrations made by hospitals of the teaching HMC, where the regional radiotherapy centre is. There were 16 registrations that should have been made by other cancer registers and 10 for which details were not available and it was not known where the previous cancer diagnosis (and treatment) had taken place.

ERRORS IN DIAGNOSIS

There were 234 cases where the 're-registration' diagnosis differed from the diagnosis in the cancer registry files; this is an error rate of 16·0% in 1460 cases studied. In Table III these differences are listed for eight broad diagnostic classifications (according to 're-registration' diagnosis) and to show (a) whether the site of cancer was entirely wrong (differences that would result in errors in the first or second digit of the *ICD* code), (b) differences that would give errors in the third digit of the *ICD* code, (c) differences that would give errors in the fourth digit of the *ICD* code and (d) second primaries that were not registered or secondaries registered in error as second primaries. Comparison of the distribution of errors by diagnosis with the distribution by diagnosis of cancer registrations indicates that there were many errors among cancers of the digestive organs (*ICD* 150-159) but that most of these (88 registrations) were errors only in the fourth digit of the *ICD* code. Examples include cancer of the fundus of the stomach (*ICD* 1518) which was registered as cancer of an unspecified part of the stomach (*ICD* 1519) and cancer of the pelvic colon (*ICD* 1533) which was registered as an unspecified part of the colon (*ICD* 1538). Of the major errors in transcription of diagnosis from hospital notes to cancer registry that led to errors in the first and second digits of the *ICD* code, there is a disproportionate excess among benign neoplasms and neoplasms of unspecified nature (*ICD* 210-239). These are chiefly papillomas of the bladder (*ICD* 2233) diagnosed several years previously which were registered (without histological confirmation) as cancers of the bladder (*ICD* 188). There were five registrations for which the hospital notes gave no record of cancer diagnoses and these five had been registered in error as haemangiomas and lymphangiomas (*ICD* 227).

TABLE II

ERRORS IN THE DATE OF REGISTRATIONS OF 1460 REGISTRATIONS SHOWING YEAR IN WHICH REGISTRATION SHOULD HAVE BEEN MADE (PERCENTAGE OF 1460 IN BRACKETS)

Site of Cancer (*ICD* code)	1972	Year in Which Registration Should Have Been Made					
		1970-71	1965-69	1960-64	1959 and before	Total	
140-149	0	0	1	0	0	1	
150-159	10	10	4	3	2	29	(1·9) L
160-163	15	2	0	0	0	17	(1·2) L
170-174	17	8	12	4	6	47	(3·2)
180-189	24	13	12	2	2	53	(3·6) H
190-199	11	0	1	0	1	13	(0·9)
200-209	6	2	4	0	0	12	(0·8)
210-239	3	6	8	4	5	26	(1·8) H
All diagnoses	86(5·9)	41(2·8)	42(2·9)	13(0·9)	16(1·1)	198	(13·6)

H significantly high rate and L significantly low rate ($P < 0·01$) compared with incidence of cancer registrations in South Wales

TABLE III

ERRORS IN DIAGNOSIS BY BROAD DIAGNOSTIC GROUPING IN 1460 CANCER REGISTRATIONS (PERCENTAGE OF 1460 IN BRACKETS)

Site of Cancer (*ICD* code)	Errors and Omissions that Would Lead to Wrong Digit of *ICD* Code			Error or Omission in Registration		
	First or Second	Third	Fourth	Second, Primary, or Secondary	All	
140-149	3	1	4	0	8	(0·5) H
150-159	10	13	88	5	116	(7·9) H
160-163	2	0	5	3	10	(0·7) L
170-174	6	1	4	13	24	(1·6) L
180-189	9	5	6	7	27	(1·8)
190-199	10	11	1	0	22	(1·5)
200-209	1	6	1	2	10	(0·7)
210-239	15	0	1	1	17	(1·2) H
All sites	56(3·8)	37(2·5)	110(7·5)	31(2·1)	234	(16·0)

H significantly high rate and L significantly low rate ($P < 0{\cdot}01$) compared with incidence of cancer registrations in South Wales

ERRORS AND OMISSIONS IN OCCUPATION

These have been enumerated under two groupings, each with two sub-groups. The first group comprises those registrations for which the occupations were either (a) completely absent from the hospital notes or (b) effectively absent from the hospital notes, as retired (with no pre-retirement occupation) or 'housewife' (with no husband's occupation). In the second group are registrations where the errors or omissions are in transcription of information from the hospital notes to the cancer registry and this group was similarly subdivided into (c) omissions and (d) failures to transcribe from the hospital notes occupation and industry details, principal lifetime occupations (instead of temporary or semi-retirement occupations), husband's occupations of married women or fathers' occupations of children. These four descriptions of error and omission of occupation from cancer registration are used mutually exclusively. There were 138 hospital records (9·5% of 1460) with no occupations and 471 (32·3%) with inadequate descriptions (retired, housewife, child, etc.) and in addition 372 (25·5%) registrations without the occupational descriptions that were in hospital notes and 109 (7·5%) with occupational descriptions that were in the hospital notes incompletely or incorrectly transcribed. Thus in total there were 1090 cancer registrations (75% of the study sample) with inadequate or no descriptions of occupations. Although a few hospital notes had good documentation of past occupations including numbers of years spent in each, such information was often not entered in cancer registration. It was also very uncommon to find the occupation of husbands of married women or of fathers of children in hospital notes.

ERRORS AND OMISSIONS IN PLACE OF BIRTH

These have been enumerated under two headings; first, omission in hospital records, which necessarily lead to omissions in the cancer registry, and secondly, failures to transcribe (correctly) information available in hospital notes to the cancer registry. There were 459 (31·4%) omissions of places of birth in hospital records and 24 (1·6%) errors and 252 omissions (17·3%) in cancer registrations of places of birth that were available in hospital records. The total error and omission rate in cancer registration is clearly the sum of these (50%). Practice varied widely from hospital to hospital within HMCs and between HMCs.

ERROR OR OMISSION IN HISTOLOGICAL TYPING

There were 272 hospital records (18·6% of 1460 registrations) without histology reports or summaries of histological findings and a further 264 registrations (19·1%) with histological type either not entered or wrongly transferred from the hospital notes.

ERRORS AND OMISSIONS IN NHS NUMBER

In only 61 (4·2%) of the 1460 registrations studied were the NHS numbers entered correctly and there were in addition only 231 hospital notes with NHS numbers, that could have been transcribed to cancer registrations. The NHS numbers were omitted from 1229 hospital notes (84·2% of 1460); a very high omission rate for a patient identification number requested by cancer registries and by OPCS.

DISCUSSION

Errors affect registration data in two ways; first, in patient identification and secondly, in the

summary statistics. Cases may be registered twice if the identification data are not correct on one of the 'registrations' and this leads to over-registration. Alternatively, over-zealous attempts to make identification data 'fit' a previously registered case with similar name, address, and date of birth could lead to loss of new cases and under-registration. Some of the reported errors in name, address, and date of birth could lead to mistaken registrations, particularly in Wales where there are relatively few surnames and many variations of spellings of place names. In practice 18 duplicate registrations from within the region and nine from other regions were identified in this study. It was found that the NHS number was quite unhelpful in patient identification because it was so little used; all patient identification and record filing in hospital is by hospital number. The reasons underlying the rare use of the NHS number have been discussed and an improved format has been suggested elsewhere (West, 1973b).

The principal summary statistics obtained from cancer registrations are incidence for each site of primary cancer by age and sex, by area of residence, and by occupation. Although the gross error rates of the relevant parameters found in this study were relatively high, the net error rates as they would affect summary statistical tabulations are generally lower. Nearly half of the diagnosis errors (7·5% of 1460) would lead to misclassification only under the fourth digit of the *ICD* code. Less than one-quarter of the date of birth errors (2·5% of 1460) would render classification in the wrong (WHO) decade and little over half of the date of registration errors (8·0% of 1460) lead to 1972 registrations when registration should have been ascribed to previous years.

The error rate in cancer diagnosis in this study (16%) is higher than that reported in the principal diagnosis of hospital inpatient morbidity data (6%) (Lockwood, 1971). However, Lockwood found much higher error and omission rates of second and subsequent diagnoses (34% and 39%) and cancer registrations may arise from diagnoses other than the principal (leading to hospital admission). In a smaller study comparing diagnoses on hospital records with those on death certificates, Alderson and Meade (1967) found 14 principal diagnosis errors among 105 hospital records. These error rates (16% in South Wales cancer registrations, 6% among Scottish hospital inpatients, and 13% among Oxford inpatients) are solely 'clerical' occurring in the preparation of summaries of the full hospital notes. It is, of course, possible that wrong diagnoses are entered in hospital notes which are then 'correctly' transcribed on to cancer registration (or hospital morbidity statistics). In a study of 450 lymphomas Hakama, Fanssila, and Saxèn (1973) found that pathologists' intra- and inter-observer variations in making diagnoses from slides were 29 and 27% respectively. High rates of disagreement have been reported when diagnoses are obtained from different raw material, by different observers and/or for different purposes. In a comparison of 1216 hospital records with the corresponding death certificates the diagnoses differed in 39% to such an extent that they were coded under different groups in the *ICD* list (Alderson and Meade, 1967) and in a study of 9501 deaths 'certified' both by clinicians and pathologists the former ascribed cancer as cause in 2283 and the latter in 2378 but diagnoses differed in 40% (Heasman and Lipworth, 1966). This study has attempted to estimate only the errors that occur summarizing information available in hospital notes and transferring the summaries to the cancer registry. HAA clerks require specific training both in medical terminology and in the use of the *ICD* code to comprehend correctly the diagnoses and to transfer correctly the relevant concise summaries. Good liaison with medical, nursing, laboratory, and technical staff should make it unnecessary for a clerk to make a diagnostic summary on her own.

The 112 registrations that should have been made before the year 1972 reflects to a considerable extent the greater efficiency of case finding by the new HAA linked cancer registration procedure than by the previous independent inquiry; the number of registrations in 1972 and 1973 were some 20% higher than in previous years (Welsh Office, 1974). When the new system has been in operation for several years the number of prevalence cases registered late should be reduced and therefore the number of errors in the date of registration should also fall. The error rate in patients' addresses (sufficient to cause wrong area of residence coding) was the lowest found, probably because it is more easily comprehended and more meaningful to records clerks than some of the other items of information and because of the uncritical assessment of error.

Place of birth, occupation and histology, like the NHS number, were omitted frequently from hospital notes and from cancer registration. Although there was provision on HAA for the collection of occupational information since its introduction in 1968 it was not often entered and, when cancer registration was linked to HAA in 1972, many hospitals failed to enter all relevant

occupational information on cancer HAA forms. However, even if all occupational information available in hospital records had been transcribed there would still have been many cancer registrations (41% of the sample) with inadequate or no information, since there was none in the hospital notes. At a recent meeting of the International Agency for Research on Cancer it was recommended that intensified use be made of established tumour registries as a basis for occupational cancer research (International Agency for Research on Cancer, 1975). This necessitates more complete recording of occupational information by cancer registries. Similarly, recording of histology was new to HAA clerks after the merging of cancer registration with HAA and the probable explanation of high omission rates is that clerks had not 'learnt' how to find the relevant information in hospital notes. Lockwood (1971) found that error and omission rates in the Scottish morbidity data study were higher in medical information than in administrative information and suggested that it could be due to poor structuring of medical records making it difficult to find certain relevant information. Finally there were significant variations between HMCs and between hospitals in error and omission frequencies in particular items of information and in practice, such as in use of the NHS number and entry of only first forename. However, there were no significant associations between error and omission frequencies in one parameter with those of another: thus if a hospital had a higher than average error rate in diagnosis it did not necessarily have higher than average error rates in name, date of birth, or occupation.

A study of transcription errors serves two useful purposes: first, in quality control and secondly, in determining an estimate of the error in cancer registry data. The fact that a study has been carried out and the feedback of its findings to those concerned with the collection of the raw data should help to reduce the more careless and unnecessary errors and omissions. Cancer registers may be used to identify particular groups of patients for epidemiological or research investigations, when minor errors in detail, which could lead to misidentification of patients, may cause considerable inconvenience. Errors in cancer registration data affect also comparative statistics in inter-regional, international or inter-occupational comparisons. However, errors that occur in the transfer of information from hospital notes to the cancer registry (as investigated in this study) or in coding this information may be few when compared with differences in description of certain classifications, particularly in international comparisons. Even within one registry there may be considerable variability in diagnostic or histological classification: for example in the Finnish study 22% of a sample of 405 lymphomas were subsequently retyped as non-malignant (Hakama *et al.*, 1973). Furthermore, despite the guidance of the *ICD*, international comparisons are often difficult to interpret because of differences in language and differences in race. Cancer registries should aim for a high standard of accuracy in data collection but at the same time it should be remembered by users of cancer statistics that significant differences in terminology or in diagnostic classification may be concealed within apparently well abstracted data.

I am grateful to the Welsh Hospital Board for making changes in their system of cancer registration and encouraging a critical study of the system, to Mrs C. Lester for her detailed analysis of 1460 hospital records and cancer registration summaries and to the hospital records departments for their co-operation.

Requests for reprints: R. R. West, MA, PhD, Department of Community Medicine, Welsh National School of Medicine, Heath Park, Cardiff.

REFERENCES

ALDERSON, M. R. and MEADE, T. W. (1967). Accuracy of diagnosis on death certificates compared with that in hospital records. *Brit J. prev. soc. Med.*, **21**, 22.

HAKAMA, M., FANSSILA, K., and SAXÈN, E. (1973). Reliability of histopathological diagnosis of malignant lymphoma. *Ann. clin. Res.* **5**, 104.

HANSLUWKA, H. (1975). In *Proceedings of the Cancer Registries and Occupational Cancer Meeting, Lyon.* International Agency for Research on Cancer, Lyon.

HEASMAN, M. A., LIDDELL, F. D. K., and REID, D. D. (1958). The accuracy of occupational vital statistics. *Brit. J. industr. Med.*, **15**, 141.

—— and LIPWORTH, L. (1966). Accuracy of certification of cause of death. *General Register Office Studies on Medical and Population Subjects, No. 20.*

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER (1975). Summary statement. In *Proceedings of the Cancer Registries and Occupational Cancer Meeting, Lyon.* International Agency for Research on Cancer, Lyon.

LOCKWOOD, E. (1971). Accuracy of Scottish hospital morbidity data. *Brit. J. prev. soc. Med.*, **25**, 76.

WELSH HOSPITAL BOARD (1974). *Cancer Registration in South Wales 1964-68, Cardiff.*

WELSH OFFICE (1974). *Health and Personal Social Services Statistics for Wales, No. 1.*

WEST, R. R. (1973a). Cancer registration by means of hospital activities analysis. *Hospital*, **69**, 372.

—— (1973b). Reorganization of health services and the NHS number. *Brit. med. J.*, **4**, 674.