How complete and accurate are cancer registrations notified by the National Health Service Central Register for England and Wales?

H O Dickinson, J A Salotti, P J Birch, M M Reid, A Malcolm, L Parker

Abstract

Study objective—To assess the completeness and accuracy of notification of cancers by the National Health Service Central Register (NHSCR) for England and Wales.

Design—Comparison of 720 cancer registrations ascertained from NHSCR up to May 1999 with those ascertained for the same cohort from six other sources and a pathology review of the NHSCR cancer registrations.

Participants—People born in Cumbria, north west England, 1950–89, and diagnosed with cancer throughout the UK, 1971–1989.

Main results—Cancer diagnoses notified by NHSCR differed substantially from those determined by this pathology review for 47 of the 688 notified cases reviewed (7%; 95% CI 5%, 9%). Over one third of these discrepancies were attributable to failures in data capture or coding by the cancer registration system and almost half to changes in diagnosis; 26 of the 47 discrepant cases were reclassified as non-malignant and 21 as malignancies but with a substantially different diagnosis. The 694 confirmed malignancies represented 94% (95% CI 92%, 95%) of the 740 cancers ascertained from all sources.

Conclusions—It is estimated that the cancer registration system missed at least 10% (95% CI 6%, 15%) of all incident cases of malignant disease. Without additional ascertainment from multiple sources and diagnostic review, it would be incautious to use NHSCR cancer registrations as the sole basis of an epidemiological study.

(J Epidemiol Community Health 2001;55:414–422)

A voluntary system of central registration of cancers has been in operation throughout England and Wales since 1971.^{1,2} Regional cancer registries ascertain details of cancer diagnoses from hospitals and other sources in their areas and forward these to the National Cancer Intelligence Centre (formerly the National Cancer Registration Bureau), which is part of the Office for National Statistics. After validation, registrations are forwarded to the National Health Service Central Register (NHSCR), which is also part of the Office for National Statistics. NHSCR is the main source for notification to epidemiological researchers of cancers diagnosed throughout England and Wales. However, there has been little evaluation of the accuracy of the registered diagnoses

and the most recent assessment of their ascertainment was restricted to children.³ This study investigates the completeness and accuracy of cancer registrations notified through NHSCR for a cohort born between 1950 and 1989 and diagnosed with cancer between 1971 and 1989, covering the age range 0–39 years.

For convenience, we refer to cancers registered within the national cancer registration system and notified to us through NHSCR as “NHSCR cancer registrations”, although NHSCR is only the final link in a chain of organisations.

Methods

THE CUMBRIAN BIRTHS DATABASE

This contains birth registration details of all 287 013 live births to mothers living in Cumbria, north west England, between 1950 and 1993 inclusive and has been described in detail elsewhere.^{4,5}

ASCERTAINMENT OF CANCERS

We ascertained cancers for the cohort from six sources in addition to NHSCR cancer registrations (see table 1).

NHSCR, the primary source of notifications, was formed in 1939 to record all residents of England and Wales in registration books containing a single line entry for each individual. These records were continually updated with all births up to 1990.⁶ Each book corresponds to a particular birth registration district and time period and is labelled with an alphanumeric code, which until 1996 formed the first part of an individual's National Health Service number. From 1971 to 1990 cancer registrations notified to NHSCR were recorded on the relevant registry entry. Death registrations were similarly linked to the relevant entry. Hence, knowing the codes corresponding to the registration districts for Cumbria for 1950–1990, NHSCR were able to scan the registry books manually and to identify individuals born in Cumbria who had developed cancer or died. These scans were carried out in 1992 and repeated in December 1995. Although NHSCR does not cover residents in Scotland, their staff liaised with their Scottish counterpart to identify events for those who had moved to Scotland and to forward death certificates and cancer registrations for them. Deaths in Northern Ireland were notified by NHSCR and certificates obtained from the General Registry Office in Belfast.

In 1991, NHSCR transferred data from the Family Health Service Authorities for those alive to the new, computerised Centralised Health Registration Information System, which largely replicates the manual system and

Department of Child Health, University of Newcastle, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne NE1 4LP, UK
H O Dickinson
J A Salotti
L Parker

Department of Histopathology, North Tyneside General Hospital, North Shields
P J Birch

Department of Haematology, Royal Victoria Infirmary, Newcastle upon Tyne
M M Reid

Department of Histopathology, Royal Victoria Infirmary
A Malcolm

Correspondence to:
Dr Dickinson
(heather.dickinson@ncl.ac.uk)

Accepted for publication
14 January 2001

Table 1 Sources of notification for cancers diagnosed 1971–89

Source of notification	Year when registrations were obtained	Age range (y) of patients registered	Cancers notified by this source	
			Total number	Not notified by NHSCR number (%)
Primary source:				
NHSCR cancer registrations:				
NHSCR registry books	1992, 1995	All ages	694*	—
Centralised Health Registration Information System	1995, 1999			
Additional sources:				
1 Northern Region Young Persons' Malignant Disease Registry	1997	0–24	336	34 (10)
2 Manchester Children's Tumour Registry	1992	0–14	11	1 (9)
3 National Registry of Childhood Tumours and Oxford Survey of Childhood Cancers†	1993/94	0–14	183	18 (10)
4 Northern Region Cancer Bureau	1993	All ages‡	288	18 (6)
5 North Western Regional Cancer Registry	1994	All ages‡	22	1 (5)
6 NHSCR death registrations:				
NHSCR registry books	1992, 1995	All ages	352	14 (4)
Centralised Health Registration Information System	1995, 1999			
Total notified by all sources			740	46 (6)

*This includes only cases confirmed on review to be malignant. NHSCR notified us of 26 further cases with morphology codes indicating malignancy that were reclassified on review as non-malignant. †Registrations of cancers for Cumbrian born children, recorded by Oxford Survey of Childhood Cancers, were obtained via National Registry of Childhood Tumours. ‡Only cases over age 15 years and for whom place of birth was known and identified to be Cumbria were notified to us.

which has recorded births, cancers and deaths registered since then. An updated list of relevant cancers and deaths for the cohort, occurring up to the end of 1993, was obtained from the Centralised Health Registration Information System in December 1995 and, again, in May 1999.

The number of cancer registrations per year received from NHSCR for the cohort reached a peak in 1989 and then fell. Hence, as of May 1999, it appeared that cancers recorded by NHSCR could not be regarded as complete beyond 1989, probably because of backlogs in transfer of registrations from regional registries to the National Cancer Intelligence Centre and hence to NHSCR.^{7,8} Therefore the study was restricted to cases diagnosed between 1971 and 1989.

Individuals in the cohort who died of cancer were also identified by scrutiny of death registrations received from NHSCR. In addition, cancer registrations for the cohort were ascertained directly from the two regional cancer registries in the north and north west of England and from regional and national children's cancer registries (see table 1).

PATHOLOGY REVIEW OF CANCER DIAGNOSES

An attempt was made to review the diagnoses of cases from a biological specimen and/or a pathology report or a postmortem report or clinical records, every effort being made to review the diagnosis from the earliest possible source in this list. When such a review was not possible, the diagnostic codes supplied by NHSCR were checked for consistency with any information from the cancer and/or death registration.

Manchester Children's Tumour Registry has centrally reviewed 96% of their registrations^{9,10} and Northern Region Young Persons' Malignant Disease Registry has centrally reviewed 81%, including re-examination of biological material for 91% of the 0–15 year age group and 70% of the 15–24 year age group for cases registered since 1971 (personal communication: Mr S Cotterill, 1999). Hence not all cases recorded by these registries were further reviewed.

The procedure used to ascertain material for review is summarised in figure 1. Solid tumours were reviewed by histopathologists (PJB, AM) and leukaemias by a haematologist (MMR).

As the different sources of ascertainment and review used different coding systems, diagnoses were translated to ICD-O-2 codes if they were originally in other coding systems (see appendix).

Cancer diagnoses in children and young people up to age 25 years were grouped into 13 major diagnostic groups following a standard classification,¹¹ updated to ICD-O-2 (personal communication: Professor J Birch and Dr V Blair, 1995), which is very similar to the International Classification of Childhood Cancers.¹² As the Birch and Marsden classification refers only to children's cancers, it was supplemented by additional groups to include adult cancers: malignancies of the gastrointestinal tract, breast, lung, cervix and other malignancies. The current audit was restricted to first primary tumours and excluded diagnoses with a morphology behaviour code of 1 or 2, indicating non-malignancy, and also excluded registrations consistent with diagnosis of cervical intraepithelial neoplasia (CIN).*

If details of the cancer had been ascertained from more than one source, the diagnostic codes and the date of diagnosis used for comparison with NHSCR were taken to be those supplied by the first available source (the review source) on the following list: the pathology review, Northern Region Young Persons' Malignant Disease Registry, Manchester Children's Tumour Registry, National Registry Childhood Tumours/Oxford Survey of Childhood Cancers, Northern Region Cancer Bureau, North Western Regional Cancer Registry, death registrations.

Finally, the NHSCR and review diagnoses were compared to see whether the site and morphology agreed and, if not, whether they were in the same diagnostic group. If they fell

* Assumed to be those with a site specified as cervix uteri, or a site specified as uterus with a morphology code that indicated a non-malignancy—1363 cases in total.

into different groups, the reviewing pathologist identified the reason for the discrepancy.

Ethical approval for the study was given by West Cumbria, East Cumbria, South Cumbria, Newcastle and Manchester Health Authority Ethical Committees.

STATISTICAL METHODS

Logistic regression was used to assess trends in accuracy of diagnosis and completeness of ascertainment with age at diagnosis and year of diagnosis.¹³ As the cohort comprised those born in Cumbria, during the period 1950 to 1989, the number of older people in the cohort increased steadily with time (see fig 2A), so age

and year of diagnosis were inevitably correlated. Therefore the trends with each of these factors was assessed after allowing for the other.

The numbers of cancers not notified by NHSCR or any other source were estimated separately for those aged 0–14 years, 15–24 years and over 25 years at diagnosis, assuming notification by NHSCR was independent of other sources. The number not notified by either NHSCR or other sources was estimated by simple proportionality, using two by two contingency tables.¹⁴

Although other proportions reported are exact (rather than estimated) for the cohort

Figure 1 Flow chart showing procedure used to carry out the diagnostic review.

Figure 2 Population at risk and numbers of cancer registrations ascertained from all sources, by age group at diagnosis and year of diagnosis.

studied, exact binomial confidence intervals (CI) are presented in the discussion to indicate the range of likely values they might take if applied to a different population.

Analysis was carried out using the statistical package Stata.¹⁵

Results

The number of children and young adults in the cohort (those born in Cumbria, 1950–89) remained roughly constant during 1971–89. However, the number of adults (over 25 years) increased steadily from 1975 onwards (see fig 2A). This pattern is reflected in the number of cancer registrations (see fig 2B, C, D).

NHSCR notified us of 720 cases diagnosed during 1971–89, 32 of which did not have specified site or morphology codes or dates of diagnosis, but were confirmed to be malignant.

ACCURACY OF DIAGNOSTIC CODES

The method of review of the remaining 688 cases whose NHSCR site and morphology codes indicated malignancy is summarised in table 2. While 206 (30%) cases were diagnosed in three hospitals in Cumbria, the remaining 482 cases were diagnosed in 113 hospitals throughout the United Kingdom, of which 105 forwarded material for review, 91 sending biological specimens. For three cases, there was

insufficient information either to confirm or refute the NHSCR diagnosis, so this was assumed to be correct.

Table 3 shows the level of agreement between the NHSCR and review diagnostic codes and the reasons for disagreement. These codes were in complete agreement for over one third of the cases (38%), in broad agreement for over half the cases (55%) and in different diagnostic groups for 47 (7%) cases, including 26 (4%) that were reclassified on review as non-malignant. The proportion of disagreements was highest in the age group 15–24 years (p for heterogeneity = 0.02, see table 4A) and was non-significantly lower in more recent years (p for trend = 0.08, see table 4B).

Over one third of the disagreements were attributable to failures in data capture and coding by the cancer registration system, (see table 3). These were mainly the supply of codes for diagnoses that had not been confirmed by histology, or codes that did not correspond to the diagnosis on the cancer registration form.

Five of the disagreements were attributable to the reviewing pathologist giving a more specific (two cases) or a less specific (two cases) diagnosis than NHSCR or an unspecific diagnosis (one case).

Almost half of the disagreements were attributable to changes in diagnosis (see table 3). Although three cases were reclassified on the basis of histopathology techniques that were not widely available at the time of the original diagnosis, most were reclassified on the basis of standard techniques (see table 3). There was no evidence that the proportion of changes in diagnosis (3% of all notifications) varied over the time period considered.

Table 5 shows the level of agreement between the NHSCR and review diagnostic codes by diagnostic group. The main group of concern was lymphoma, of which 13 cases

Table 2 Method of review of cancers notified by NHSCR and diagnosed between 1971 and 1989

Method of review	Number (%)
Biopsy*	413 (60)
Pathology report*	68 (10)
Postmortem report*	13 (2)
Clinical records only	128 (19)
Information provided by cancer registry	66 (10)
Total	688 (100)

*With or without other records.

KEY POINTS

- We estimate that NHSCR notified us of 90% (95%CI 85%, 94%) of all incident cancers.
- Completeness of ascertainment of cancers from NHSCR alone may not be adequate for epidemiological studies.
- Registrations should be sought from several additional sources.
- Likewise NHSCR site and morphology codes should be confirmed where possible.

(10%) were reclassified, three as leukaemia, nine as non-malignant and, for one case, the reviewing pathologist found no evidence from the postmortem report of a malignancy but insufficient evidence to exclude it.

ACCURACY OF DIAGNOSIS DATE

For the 555 cases notified by both NHSCR and another source, which were confirmed as malignant, the dates of diagnosis recorded by NHSCR and the review source were compared: 94% of these were within six weeks of each other and 99% within a year, but for 2%

the calendar year of diagnosis was different. For the six for which the dates of diagnoses differed by more than a year, all information was rechecked: for two cases the NHSCR date of diagnosis was correct and the review source was incorrect, for one case NHSCR had used the date of death rather than the date of diagnosis and for three cases the review source (Northern Region Young Persons' Malignant Disease Registry) was correct and had an earlier date of diagnosis than NHSCR.

COMPLETENESS OF ASCERTAINMENT

The 694 cases notified to us by NHSCR and confirmed as malignant represented 94% of the total of 740 confirmed malignancies notified from all sources, (see table 1).

For those aged 0–14 years, we approached three specialist children's cancer registries—two regional and one national; for those aged 15–24 years we approached one specialist regional registry for young people and two regional cancer registries; for those aged over 25 years we approached two regional cancer registries (see table 1). Therefore we expected ascertainment in the younger age groups to be better. This was confirmed by the significantly higher percentage of cases known to us that

Table 3 Level of agreement between site and morphology codes assigned by NHSCR and diagnostic review and reasons for disagreement

Level of agreement between diagnostic codes of NHSCR and review and reasons for disagreement	Review diagnosis		Total number of cases Number (%)†
	malignant Number (%)†	non-malignant Number (%)†	
Complete agreement in site and morphology codes	263 (38)		263 (38)
Agreement in diagnostic group* but not in site and morphology codes	378 (55)		378 (55)
Disagreement in diagnostic group*	21 (3)	26 (4)	47 (7)
(a) Disagreements due to failures in data capture and coding by national cancer registration system	7	12	19
Unconfirmed morphology code supplied, eg 6xxx or 7xxx	1	7	8
Coding did not correspond to diagnosis on cancer registration	4	4	8
Wrong person flagged	0	1	1
Metastatic—reviewing pathologist could not confirm the site code which may have indicated the site of a biopsy rather than the primary cancer	2	0	2
(b) Reviewing pathologist gave a more or less specific diagnosis than NHSCR	5	0	5
Review diagnosis less specific	2	0	2
Review diagnosis more specific	2	0	2
Evidence not consistent with malignancy but insufficient evidence to diagnose as non-malignant	1	0	1
(c) Disagreements due to different diagnosis	9	14	23
Review diagnosis based on standard techniques	6	14	20
Review diagnosis based on immunocytochemistry	3	0	3
Total	662 (96)	26 (4)	688 (100)

*Diagnostic groups were assigned using a modified Birch and Marsden classification, see text. †Percentages are relative to the total of 688 malignancies notified by NHSCR.

Table 4 Trends in (i) accuracy and (ii) completeness of notifications from NHSCR by (a) age at diagnosis and (b) time period of diagnosis

	(a) Age (y) at diagnosis					
	0–14 Number (%)	15–24 Number (%)	over 25 Number (%)	all ages Number (%)	p for heterogeneity between age groups (adjusting for year of diagnosis)	
(i) Cases with NHSCR and review diagnoses in different groups*	8 (5)	25 (11)	14 (5)	47 (7)	0.02	
(ii) Total number of cancers ascertained from all sources	195 —	233 —	312 —	740 —	—	
Cancers missed by NHSCR†	20 (10)	15 (6)	11 (4)	46 (6)	0.02	
	(b) Time period of diagnosis					
	1971–74 Number (%)	1975–79 Number (%)	1980–84 Number (%)	1985–89 Number (%)	1971–89 Number (%)	p for trend with year of diagnosis (adjusting for age group at diagnosis)
(i) Cases with NHSCR and review diagnoses in different groups*	9 (13)	9 (8)	11 (6)	18 (6)	47 (7)	0.08
(ii) Total number of cancers ascertained from all sources	74 —	128 —	206 —	332 —	740 —	—
Cancers missed by NHSCR†	5 (7)	8 (6)	14 (7)	19 (6)	46 (6)	0.83

*Percentages are relative to the number of cases in each category of age group or time period that were notified by NHSCR with morphology codes indicating malignancy. †Percentages are relative to number of malignancies in each category of age group or time period that were ascertained from all sources.

Table 5 Summary of review diagnoses compared to diagnoses indicated by NHSCR site and morphology codes. Disagreements were due to (a) failures in data capture and coding by national cancer registration system; (b) the reviewing pathologist giving a more or less specific diagnosis than national cancer registration system; (c) the reviewing pathologist giving a different diagnosis from national cancer registration system

		Number of cases for which diagnostic group of NHSCR and review is									
		same		different							
		Review diagnosis									
Diagnostic group corresponding to NHSCR site and morphology codes	Total number of cases	malignant				non-malignant				Total number of changed diagnoses	
		a	b	c	All	a	b	c	All		
Leukaemias	87	86	—	—	—	0	—	—	1	1	1
Lymphomas and other lymphoreticular neoplasms	126	113	—	1	3	4	3	—	6	9	13
Central nervous system, intracranial and intraspinal neoplasms	74	73	—	—	—	0	1	—	—	1	1
Retinoblastoma	4	4	—	—	—	0	—	—	—	0	0
Adrenal and other peripheral neural tumours	12	12	—	—	—	0	—	—	—	0	0
Renal tumours	16	15	—	—	1	1	—	—	—	0	1
Hepatic tumours	5	4	1	—	—	1	—	—	—	0	1
Bone tumours	30	27	1	—	1	2	1	—	—	1	3
Soft tissue sarcoma	35	31	—	—	2	2	1	—	1	2	4
Germ cell, trophoblastic and other gonadal tumours	87	85	—	1	—	1	1	—	—	1	2
Skin tumours	75	69	—	—	—	0	3	—	3	6	6
Tumours of gastrointestinal tract	24	24	—	—	—	0	—	—	—	0	0
Breast tumours	59	57	2	—	—	2	—	—	—	0	2
Lung tumours	12	10	—	1	1	2	—	—	—	0	2
Specified malignant neoplasms NEC	42	31	3	2	1	6	2	—	3	5	11
All tumours	688	641	7	5	9	21 (3%)	12	0	14	26 (4%)	47 (7%)

were not notified by NHSCR in the younger age groups (see table 4A: 10%, 6% and 4% in the age groups 0–14, 15–24 and over 25 years respectively, p for heterogeneity=0.02). This under-notification by NHSCR was more marked for cases who had not died (NHSCR failed to notify 16%, 10% and 3% of such cases in the respective age groups, p for heterogeneity =0.006). After stratifying by age group there was little variation in the level of notification over the time period considered ($p=0.83$, see table 4B).

There must inevitably be other cancers, not ascertained by us from any source. If ascertainment from NHSCR cancer registrations and the other sources listed in table 1 were independent, the estimated number of such cancers would be 0.0, 1.9 and 4.4, in those aged 0–14, 15–24 and over 25 years at diagnosis, respectively. Adding these estimates of cancers missed by all sources to the numbers of cancers missed by NHSCR but known from other sources (see above) implies that NHSCR missed a total of 10% (95%CI 6%, 15%), 7% (95%CI 4%, 11%) and 5% (95%CI 3%, 8%) of actual cancers in these age groups and 7% (95%CI 5%, 9%) of cancers overall.

Of the 26 cancer registrations in people over 15 years not notified by NHSCR, 19 (73%) were notified to us by regional cancer registries (Northern Region Cancer Bureau or North Western Regional Cancer Registry), which notify cancers to the National Cancer Intelligence Centre and hence would be expected to be recorded by NHSCR; this proportion did not change significantly over time.

Discussion

A national system of cancer registration, such as National Cancer Intelligence Centre, is a prerequisite for monitoring geographical, social and occupational trends in cancer incidence and for the development and evaluation of policies on prevention and screening of cancer.^{16 17} However, many epidemiological studies

require the linkage of cancer details to individuals. NHSCR provides such a facility for England and Wales, aiding the selection of samples from the entire population for inclusion in cohort or case-control studies. Individuals can be “flagged” for retrospective studies, researchers being informed if their study members have developed cancer, died or emigrated whereas, for prospective studies, researchers are informed if and when these events occur.⁶ Hence, many epidemiological investigations rely on the timeliness, completeness and accuracy of information passed to researchers by NHSCR,¹⁸ although for some studies, data obtained directly from regional registries may be more appropriate.¹⁹

Although the need for quality assurance of cancer registration data is recognised,^{20 21} few systematic audits have been carried out to date of the completeness and accuracy of cancer data recorded on NHSCR, (see table 6). Although the cohort considered in this study does not constitute a random sample of cancer cases in England and Wales, being biased towards those diagnosed at a younger age and in more recent years, it extends the findings of previous studies.

ACCURACY OF DIAGNOSTIC CODES

There were substantial disagreements between the NHSCR diagnoses and the review diagnoses in 7% (95%CI 5%, 9%) of cases, approximately half of these disagreements resulting in reclassification as non-malignant on review. These disagreements could have implications for epidemiological studies, either of cancers within specific diagnostic groups or of all malignancies.

Over one third of the disagreements were attributable to failures in data capture by the national cancer registration system. In 3% (95%CI 2%, 5%) of cases, there were genuine differences between the original diagnosis and the review diagnosis. The main areas of concern were lymphomas, of which 10%

Table 6 Studies of completeness of ascertainment of cancers notified by NHSCR

Study group	Cancers ascertained from NHSCR and the sources below	Outcome studied (diagnosis or death)	Time period of diagnoses	Date of study	Number of cases ascertained	% notified to researchers by NHSCR	Reference
People born in Cumbria, 1950-89	See table 1	diagnosis	1971-89	May 1999	740	94	Present study
Children, aged 0-14 years, resident in England and Wales	National Registry of Childhood Tumours	diagnosis	1971-84	Dec 1989	7379	92	Hawkins and Swerdlow, 1992 ³
Children, aged 0-19 years, resident in England and Wales	National Registry of Childhood Tumours	death	1953-88	1989	614	96	Hawkins and Swerdlow, 1992 ³
Women in Oxford Family Planning Association contraceptive study	Self reported morbidity	diagnosis	1968-85	June 1987	325	67	Villard-Mackintosh et al, 1988 ²³
Women with breast cancer	Clinical trial, self reported morbidity	diagnosis	1973-82	May 1985	50	44	Hunt and Coleman, 1987 ²⁴
Women with ovarian cancer	Self reported morbidity	diagnosis	1986-93	July 1997	49	78	MacDonald et al, 2000 ²⁷

(95%CI 6%, 17%) were given a different diagnosis on review. Histopathology is the "gold standard" in the diagnosis and typing of the vast majority of tumours. However, its interpretation is to a certain extent subjective, relying on the experience and expertise of the pathologist and hence there will inevitably be occasional differences of opinion. It is worth noting also that better diagnostic criteria may result in changes to a diagnosis made many years ago. Only three disagreements in diagnosis (0.4% of all cancers) resulted from the application of advanced diagnostic techniques available to the pathologists reviewing the cases but not to those making the original diagnoses. However, in some instances the reviewing pathologists had the benefit of hindsight in that they had access to follow up records of the patients.

It is important to note that the diagnosis on which treatment was based was not necessarily the diagnosis reported through NHSCR and we do not know if the discrepancies we observed had any implications for the individual patients concerned.

ACCURACY OF DIAGNOSIS DATE

Agreement on date of diagnosis was generally high, with dates for fewer than 1% of the cancers notified through NHSCR discrepant by over one year.

COMPLETENESS OF NOTIFICATION

In this study, 94% of cancers ascertained during 1971-89 were notified to us by NHSCR by May 1999. This level of ascertainment is compared with other studies in table 6. There is a known delay between diagnosis and recording of a cancer on NHSCR,^{3, 22} which would account for the low level of notification observed in studies carried out within three years of diagnosis.²³⁻²⁷ Our study had a minimum period of 10 years between diagnosis and assessment of completeness of NHSCR notifications, longer than that reported in the study of Hawkins and Swerdlow, 1992,³ which may account for its higher level of notification.

It is of concern that 19 (73%) of the 26 cancers not notified by NHSCR in those over 15 years were known to the relevant regional registry (Northern Region Cancer Bureau or North Western Regional Cancer Registry) indicating difficulties at some point in the flow of information from regional registries to the National Cancer Intelligence Centre and hence to NHSCR and researchers.²² Some records may not have been sent by the regional registries to the National Cancer Intelligence Centre, some records may not have been sent from National Cancer Intelligence Centre to NHSCR at the time of our study or may have gone astray between them, some flagged cases may not have been notified to researchers and some cases may have been wrongly flagged.

It is clear that, as noted by other researchers,²⁸ ascertainment of cancers is likely to be inadequate for epidemiological studies unless registrations are sought from several sources: six of the seven sources that we approached

notified details of cancers unknown to all other registries.

There were almost certainly some cancers not registered by any of the sources we approached. The numbers of such cancers were estimated by two methods. Firstly, assuming independence of ascertainment by NHSCR and other sources, the proportion missed was estimated to be 7%. This is likely to be an underestimate, as the assumption of independence of sources is not valid, as NHSCR is informed of cancers registered by the regional registries, which exchange information with the specialist children's and young persons' registries. Secondly, the apparently worse notification by NHSCR of children's cancers is almost certainly an artefact attributable to our more thorough ascertainment of childhood cancers from regional and national specialist children's cancer registries. Hence the proportion missed in all age groups is likely to be at least 10% (95% CI 6%, 15%), as in the youngest age group, despite a generous interval of at least 10 years between diagnosis and notification. In this series of 694 cancers diagnosed at all ages and notified by NHSCR, 10% missed would represent 77 cases.

It is clear that overall case ascertainment was superior for the younger age group, as a consequence of the availability of specialist children's cancer registries, which are an important resource for epidemiological studies of these rare tumours.

REPRESENTATIVENESS OF STUDY

There is varying under-ascertainment of cases by regional cancer registries.² About half the cases in this study were diagnosed within the area covered by the Northern Region Cancer Bureau, which was one of the few regional registries in the United Kingdom whose data were not included in *Cancer Incidence in Five Continents*,²⁹ which has stringent data quality criteria. Hence it is possible that the level of notification might have been higher if the study had been based on a different geographical cohort. On the other hand, there are many steps in the supply of information where cases may have been lost and NHSCR failed to notify us of cases known to the Northern Region Cancer Bureau. It is therefore unclear whether geographical bias has affected the level of notification.

Information for diagnostic review was obtained from over 100 hospitals throughout the UK in addition to the two major regional referral centres for children's cancers (Newcastle in the northern region and Manchester in the north western region). About half of the disagreements in diagnosis were for cases diagnosed outside the area covered by the Northern Region Cancer Bureau, corresponding roughly to the distribution of diagnoses. Therefore the 3% disagreement in diagnosis is likely to be typical of the country as a whole.

Although the age distribution in our cohort changed over time, it consisted predominantly of children and young people and included no one aged over 40 years (see fig 2). This was reflected in the distribution of tumours (see

table 5), in particular the high proportion of lymphomas (18% of all tumours but 28% of those reclassified). While the incidence of cancer increases dramatically with age, the proportion of lymphomas declines; hence the rate of disagreement for a sample with an age distribution representative of all cancer cases might be lower than that found in our cohort.

In conclusion, we estimate that NHSCR notified us of 90% (95%CI 85%, 94%) of all incident cancers and that this is likely to be typical of notification of cancers diagnosed under 40 years of age throughout England and Wales. Among the cancers notified by NHSCR as malignant, 7% (95%CI 5%, 9%) had diagnostic codes that differed substantially from those assigned by our pathology review, 4% (95%CI 2%, 5%) being reclassified as non-malignant.

Completeness of ascertainment of cancers from NHSCR alone may not be adequate for epidemiological studies and registrations should be sought from several additional sources where possible. Likewise, researchers should be aware that a small fraction of site and morphology codes notified by NHSCR may not be accurate and diagnostic review should be attempted.

The National Cancer Intelligence Centre is the largest national cancer registry in the world. By linking this information on cancers to individuals and combining it with information on deaths and emigrations, NHSCR provides a uniquely powerful resource for medical research. It is therefore critical that checks on its accuracy and completeness are assured through regular audit and administrative excellence. Adequate funding is clearly of national importance, not only for NHSCR but for the entire cancer registration infrastructure, including primary data suppliers in hospitals and pathology laboratories, regional registries and the National Cancer Intelligence Centre.

We thank the following for extracting details of Cumbrian born cancer cases: Mrs Lorna More at the Northern Region Young Persons' Malignant Disease Registry, Professor Jill Birch and staff at the Manchester Children's Tumour Registry, Dr Gerald Draper and staff at the National Registry of Childhood Tumours, Dr Tom Sorahan and Dr Estelle Gilman at the University of Birmingham, Mr John Stephenson and Mrs Carol McCarthy at the Northern Regional Cancer Bureau and Dr Tony Moran of the North Western Regional Cancer Registry. We thank pathologists, haematologists and staff at hospitals throughout the UK for facilitating the review process. We thank Dr M Quinn, Director of the National Cancer Intelligence Centre, for constructive comments on the manuscript. We thank the Office of Population Censuses and Surveys for providing us with birth registrations, Mr David Harris and staff at the National Health Service Central Register for providing us with

cancer, death and embarkation data. We thank the following staff of the North of England Children's Cancer Research Unit, University of Newcastle: Mr Julian Smith for systems analysis and programming, Mrs Katharine Kirton for secretarial assistance and Mrs Christine Kinsella for extracting records of cancer cases.

Funding: we thank Westlakes Research Institute and the North of England Children's Cancer Research Fund for contributing to the funding of the project.

Conflicts of interest: none.

Appendix

As the different sources of ascertainment and review used different coding systems (see above), diagnoses were translated to ICD-O-2 codes if they were originally in other coding systems. All unique combinations of site and morphology codes were extracted. ICD-8, ICD-9 and SNOMED codes were translated to ICD-O-2 codes using the relevant manuals and ICD-O-1 codes were converted to ICD-O-2 using a computer program, CONVERT.³⁰ There were various difficulties with translations of the coding systems. ICD-8 does not specify morphology codes, but from 1971 onwards the Office for National Statistics used an in house system for morphology codes based on MOTNAC.³¹ These were four digit codes, with the fourth digit usually indicating the behaviour (benign, uncertain, in situ, malignant, secondary) of the neoplasm, for example the ICD-9 code 8070/3 was coded as 8073. Also the ICD-9 morphology codes supplied to the Office for National Statistics by some regional cancer bureaux did not have a fifth digit indicating the behaviour of the tumour as the site code usually indicates this; in such cases a behaviour code was added during translation. In some cases the ICD-9 coding system had been supplemented by use of SNOMED codes for B cell and T cell lymphomas. All histological diagnoses provided to the Office for National Statistics for years of registration between 1971 and 1992 whose morphology was unconfirmed were indicated by codes from 6000 to 7999 (2000 less than the corresponding confirmed code). During validation, 2000 was added to each of these codes (for example 6000 became 8000) and checked against the relevant ICD classification.

Morphology codes within the SNOMED system were the same as those within ICD-O-2 except for lymphomas and a few diagnoses such as borderline malignancies, for example, mucinous cystadenoma, which has a morphology code of 8470/1 in SNOMED and 8472/3 in ICD-O-2.

Nine codes could not be assigned a unique ICD-O-2 code, mainly because the more recent ICD-O-2 classification used a more precise classification for the site of the tumour. In such cases, textual information from the relevant registration was consulted to aid in recoding.

For nine cases for which NHSCR did not supply site or type codes, the diagnosis was noted on the NHSCR card recording the tracing of the case and was coded by us.

Appendix: Coding of cancer diagnoses: coding systems used by registries and pathology review

Coding system	Registry
ICD-8*	Office for National Statistics, Northern Region Cancer Bureau, North Western Regional Cancer Registry: during 1968–78
ICD-9*	Office for National Statistics, Northern Region Cancer Bureau, North Western Regional Cancer Registry: from 1979
ICD-O-1†	Manchester Children's Tumour Registry
ICD-O-1†	National Registry of Childhood Tumours
ICD-O-1†	Northern Region Young Persons' Malignant Disease Registry
SNOMED‡	Pathology review—solid tumours
ICD-O-2‡	Pathology review—leukaemias

*ICD-8, ICD-9 Eighth (1967) and Ninth (1977) Revisions of International Classification of Diseases. WHO. †ICD-O-1, ICD-O-2 International Classification of Diseases for Oncology, editions 1 (1976) and 2 (1990). WHO. ‡SNOMED Systematized Nomenclature of Medicine, 2nd ed (1976). College of American Pathologists.

- Office of Population Censuses and Surveys. Series MB1 no 17. *Review of the national cancer registration system*. London: HM Stationery Office, 1990.
- Office for National Statistics Series MB1 no25. *1992 Cancer statistics: registrations. England and Wales*. London: HM Stationery Office, 1998.
- Hawkins MM, Swerdlow AJ. Completeness of cancer and death follow-up obtained through the National Health Service Central Register for England and Wales. *Br J Cancer* 1992;66:408–13.
- Parker L, Smith J, Dickinson H, *et al*. The creation of a database of children of workers at a nuclear facility: an exercise in record linkage. *Appl Occup Environ Hygiene* 1997;12:40–45.
- Dummer TJB, Dickinson HO, Pearce MS, *et al*. Stillbirth risk with social class and deprivation: no evidence for increasing inequality. *J Clin Epidemiol*. 2000;53:147–55.
- Botting B, Reilly H, Harris D. Use of Office of Population Censuses and Surveys records in Medical research and Clinical Audit. *Health Trends* 1995;27:4–7.
- Quinn MJ. Cancer redevelopment. *The Researcher* 1996;6:6.
- Quinn MJ. Update on progress on flagging cancers at NHSCR Southport. *The Researcher* 1999;9:10.
- Blair V, Birch JM. Patterns and temporal trends in the incidence of malignant disease in children: I. Leukaemia and lymphoma. *Eur J Cancer* 1994;30A:1490–8.

- 10 Blair V, Birch JM. Patterns and temporal trends in the incidence of malignant disease in children:II. Solid tumours of childhood. *Eur J Cancer* 1994;**30A**:1498–511.
- 11 Birch JM, Marsden HB. A classification scheme for childhood cancer. *Int J Cancer* 1987;**40**:620–4.
- 12 Kramárová E, Stiller CA, Ferlay J, et al. *International classification of childhood cancer 1996*. IARC Technical Report no 29. Lyon, France: International Agency for Research on Cancer, 1996.
- 13 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley, 1989.
- 14 Brenner H. Use and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology* 1995;**6**:42–8.
- 15 Stata Corp. *Stata Statistical Software: release 5.0*. College Station: Stata Corporation, 1997.
- 16 Dickinson HO. Cancer trends in England and Wales. *BMJ* 2000;**320**:884–5.
- 17 Coleman MP, Esteve J, Damiecki, et al. *Trends in cancer incidence and mortality. IARC Scientific Publications no 121*. Lyon, France: International Agency for Research on Cancer, 1993.
- 18 Moss SM, Coleman DA, Chamberlain J, on behalf of UK Trial of Early Detection of Breast cancer group. 16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer. *Lancet* 1999;**353**:1909–14.
- 19 Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *BMJ* 1995;**311**:1391–5.
- 20 Winyard G. An advisory committee on cancer registration has now been established *BMJ* 1996;**313**:1149.
- 21 Day NE, Davies TW. Cancer registration: integrate or disintegrate? *BMJ* 1996;**313**:896.
- 22 Quinn MJ. Mike Quinn, Director of the National Cancer Registration Bureau (NCRB) explains why cancer cases may not be notified to researchers (yet). *The Researcher* 1998;**8**:6.
- 23 Villard-Mackintosh L, Coleman MP, Vessey MP. The completeness of cancer registration in England: an assessment from the Oxford FPA contraceptive study. *Br J Cancer* 1988;**58**:507–11.
- 24 Hunt K, Coleman MP. The completeness of cancer registration in follow-up studies. A cautionary note: *Br J Cancer* 1987;**56**:357–9.
- 25 MacDonald N, Sibley K, Rosenthal A, et al. A comparison of national cancer registry and direct follow-up in the ascertainment of ovarian cancer. *Br J Cancer* 1999;**80**:1826–7.
- 26 Quinn MJ, Botting B, Foote D, et al. Registration of ovarian cancer in England and Wales. *Br J Cancer* 2000;**83**:278–9.
- 27 MacDonald N, Menon U, Jacobs I. Registration of ovarian cancer in England and Wales. *Br J Cancer* 2000;**83**:279.
- 28 McKinney PA, Ironside JW, Harkness EF, et al. Registration quality and descriptive epidemiology of childhood brain-tumours in Scotland 1975–1990. *Br J Cancer* 1994;**70**:973–9.
- 29 Parkin DM, Muir CS, Whelan SL, et al. *Cancer incidence in five continents. Volume VI IARC Scientific Publications no 120*. Lyon, France: International Agency for Research on Cancer, 1992.
- 30 Ferlay J. *ICD Conversion programs for cancer. IARC Technical Report no 21*. Lyon, France: International Agency for Research on Cancer, 1994.
- 31 American Cancer Society. *Manual of tumor nomenclature and coding*. New York: American Cancer Society, 1968.