Reliability of data of the Thames cancer registry on 673 cases of colorectal cancer: effect of the registration process

Allyson M Pollock, Neil Vickers

Abstract

Objective—To measure the reliability of data collected by the Thames cancer registry and to identify factors in the registration process affecting reliability.

Design—A retrospective study of data from the registry, including death certificate only registrations, and hospital case notes on cases of colorectal cancer diagnosed in 1983 or 1988.

Setting—Four districts in South Thames region.

Subjects—673 cases of colorectal cancer in resident patients.

Main measures—Dates of birth, diagnosis of cancer, and death; sex; tumour site; whether treatment was given; type of treatment; and district of residence.

Results—Among the 416 (62%) case notes retrieved, including 66 death certificate only registrations, full or high agreement between registry data and hospital notes was recorded for sex, district of residence, and dates of birth and death. Only 12% of cases had the same date of diagnosis, which may be due to failure of registry policy. Lower agreement rates occurred for tumour site (87%), whether treatment occurred (84%), and treatments administered (80%, 1983; 72%, 1988). 20% of surgical treatments and 37% of adjuvant therapy, radiotherapy, and chemotherapy were not recorded by the registry. Disagreements were common among death certificate only registrations. Such registrations accounted for 16(32%) disagreements over tumour site, 33(41%) major disagreements over date of diagnosis (difference > 30 days), and 47(44%) disagreements over treatment. In 65 cases the registry failed to capture all treatments carried out within the six month follow up period, 38(58%) of which were for death certificate only registrations. In 36% of death certificate only registrations the patients survived more than one year from diagnosis, indicating a failure of registry policy over retrospective follow up.

Conclusions—Registry data on district of residence; sex; dates of birth, diagnosis, and death are highly reliable, but treatment and tumour site data are less so. Lack of follow up in death certificate only registrations and failure to monitor treatments during follow up period seemed to be associated with disagreements.

Implications—Attention to these registration processes might greatly improve reliability of registry data.

Keywords: Cancer registry data, colorectal cancer, cancer registration.

Introduction

National cancer statistics for England and Wales are derived from 12 population based cancer registries.1 Their validity depends on the completeness of case ascertainment2-5 by the registries and on the completeness and accuracy of the data sources from which registrations are made.6-10

Until recently, concerns about the completeness and reliability of cancer registry data have restricted the use of the data in evaluating health care. However, recent changes to the structure of the UK National Health Service, together with the introduction of clinical audit, have stimulated new interest in the use of cancer registry data to evaluate health service delivery.11 Questions surrounding the completeness and accuracy of registry data have consequently assumed a new urgency.

The Thames cancer registry is a geographically defined register covering a population of 13-8 million residents of the four regional health authorities in south east England. Since it contributes up to a third of all data on cancer for England and Wales missing or poor quality data will have serious implications for national baseline statistics.

PROCESS OF REGISTRATION BY THE REGISTRY

Figure 1 describes the process of registration used by the Thames cancer registry and highlights the registry's policy on follow up of cases. The main sources of data for cancer registration are hospital case notes and copies of death certificates supplied by the Office of Population Censuses and Surveys (OPCS) to each registry for every person for whom cancer is mentioned who dies on its territory. Data are abstracted from the case notes at treatment centres by peripatetic registration officers trained by the registry.12-15 Data items retrieved include sociodemographic details, tumour site and stage, the first six months' planned treatment, district of residence and district of diagnosis. The registry follows up cases for the first six months after diagnosis.

The other main data source for registration is death certificate notifications. These enable
will not be available when death certificates are the primary data source. Other variables also may be inaccurate. Percy et al reported only 65% accuracy in the coding of cause of death in death certificates of cancer patients. Since confirming a date of diagnosis for a death certificate only registration is rarely possible, the date of diagnosis is always given as the date of death, although this may be incorrect. We would expect these deficiencies to undermine the reliability of cancer registry data.

**Coding agreement and errors of data entry**

Error may arise in transcribing data from the case notes to the registry proforma or as a result of differences between clinical practices and registry coding practices. For example, the Thames cancer registry defines date of diagnosis by reference to a hierarchy of clinical events (first diagnosis by histology, other diagnostic information (for example, computed tomographs), clinical opinion, death), and this may not correspond to the system used by clinicians.

This paper describes the results of a study set up to investigate retrospectively the completeness and accuracy of data of the Thames cancer registry on 673 cases of colorectal cancer, using case notes as a standard. The study differs from previous studies in that it pays special attention to the ways in which the process of registration can affect data accuracy and completeness and thereby identifies aspects of registry policy which might be amenable to intervention.

**Methods**

We performed a retrospective case note study of all cases of colorectal cancer in the Thames cancer registry diagnosed in 1983 or 1988 in patients resident in any of four districts (A–D) in the South Thames regions. In 1990 we used Thames cancer registry data to carry out an ecological analysis of all cases of colorectal cancer diagnosed from 1982 to 1988 inclusive in the South Thames regions. The registry supplied data on all cases in electromagnetic form. To investigate the underlying reasons for differences in incidence and survival we undertook a retrospective case note study in four districts (the two with the best survival and the two with the worst). This study began in May 1991. We chose to look only at cases with a diagnosis in 1983 and in 1988 (the most recent year for which we had full data). Since hospitals usually retain case notes for eight years from the date of last attendance, we could not go further back than 1983. We wanted to look at data for two years in case any significant changes in the management of colorectal cancer occurred over time. The study described here developed out of this work.

Using Thames cancer registry data as a sampling frame, we asked medical records departments in the four districts to seek and retrieve case notes. We excluded cases in which diagnosis and treatment seemed (in the registry) to have occurred outside the district of residence, which accounted for 17% of cases.

**Completeness of primary data source**

The policy of the Thames cancer registry is to follow up case notes actively for six months from the date of diagnosis. But the registry may not capture data on treatments administered after this period. It may also fail to capture data on treatment if monitoring does not take place over the whole six months.

The registry attempts to follow up all death certificate initiated registrations, since data on date of diagnosis, tumour stage, and treatment...
in each year, but we requested notes on death certificate only registrations. Case notes were retrieved only from those NHS hospitals with responsibility for NHS hospital records in each district. In districts A, B, and C this task devolved to one hospital alone; in district D it was shared by two hospitals. In total, there were five medical record sites covering six hospitals and outlying outpatient departments.

Data were retrieved by a consultant (AMP) and a senior house officer in public health medicine using a structured proforma. Eight variables were extracted from each set of case notes: dates of birth, diagnosis, and death; sex; tumour site (colon, rectum, or site unspecified); whether treatment was given (yes, no, or unknown); the type of treatment (surgery, adjuvant therapy, chemotherapy or radiotherapy, or no treatment); and district of residence. These variables are routinely collected by the registry. The category “adjuvant therapy” was used to designate surgical treatment supplemented with chemotherapy or radiotherapy, or both. “No treatment” included cases for which no treatment was listed as well as those for which it was positively indicated that no treatment had taken place. The retrieval process took 18 months and in some districts required repeat visits; telephone calls; and extra help from audit coordinators, local research assistants, and clinicians.

AGREEMENT BETWEEN CANCER REGISTRY AND CASE NOTES

Percentage agreements between the registry and case notes were recorded for each of the eight variables. Date variables were classified into four categories: absolute agreement between the registry and the case notes (that is, to the day), agreement within 30 days of each other, agreement within 90 days of each other, and agreement greater than 90 days of each other. Case notes without a date of diagnosis or a date of death were excluded. This accounts for the different denominators observed for each variable.

FACTORS ASSOCIATED WITH DISAGREEMENT

Source of registration – Some disagreements may occur because the registry used death certificates as its final source of registration. Disagreements over treatment variables, tumour site, and date of diagnosis were analysed by data source for registration – that is, case notes or death certificates.

Timing of registration – Since the registry may fail to capture some treatments because they were given after the six month follow up we also analysed cases according to the time between date of diagnosis and treatment (more or less than six months).

Survival – A higher proportion of poor survival would be expected in the death certificate only registrations, since the registry would have had less time to register these in life. To establish whether cases with a death certificate only registration with missing treatment data had shorter survival time than cases registered from case notes with missing treatment data, we cross tabulated survival time against final source of registration. Case notes without a date of diagnosis were excluded from this analysis.

Results

In all, 673 cases of colorectal cancer registered by the Thames cancer registry in 1983 and 1988 met our criteria for inclusion in the study (table 1), of which 290 were registered in 1983 and 383 in 1988. One hundred and fifty (22%) cases were death certificate only registrations, 64 in 1983 and 86 in 1988. A total of 416(62%) sets of case notes were retrieved, 117(40%) for 1983 and 299(78%) for 1988. This included 66 sets of case notes for death certificate only registrations, 12(19%) for 1983 and 54(63%) for 1988. District C failed to retrieve any cases for 1983 and retrieved only 26% for 1988. Poor retrieval was also recorded for cases diagnosed in 1983 in district A. No significant difference in case note retrieval by year of diagnosis was found in districts B and D.

AGREEMENT BETWEEN CANCER REGISTRY AND CASE NOTES

Rates of agreement for sex and district of residence approached 100% (table 2). There was strong agreement between case notes and the registry for dates of birth and death but weak agreement for date of diagnosis (table 3). Absolute agreement on date of diagnosis was only 12% although this rose to 82% when discordant results falling within 30 days of case note date of diagnosis were included. However, in 18% of case notes disagreements were greater than 90 days, and in 40 cases the date of diagnosis fell in a different year from that specified in the registry.

Rates of agreement on tumour site were 88% for both 1983 and 1988 (table 4). In 23(6%) cases no primary site could be ascertained from the case notes; 30(12%) and 19(12%) disagreements were disclosed between the case notes and the registry data for colon cancer and rectal cancer respectively. The registry misclassified 10 rectal tumours as colon tumours and 16 colon tumours as rectal tumours.

---

Table 1  Retrieval rates of cases of colorectal cancer by district of residence and year of diagnosis. Figures are numbers (percentages)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11/85(13)</td>
<td>119/132(90)</td>
<td>130/217(60)</td>
</tr>
<tr>
<td>B</td>
<td>62/73(85)</td>
<td>95/102(93)</td>
<td>157/175(90)</td>
</tr>
<tr>
<td>C</td>
<td>0/97</td>
<td>37/76(49)</td>
<td>37/143(26)</td>
</tr>
<tr>
<td>D</td>
<td>44/65(68)</td>
<td>48/73(66)</td>
<td>92/138(67)</td>
</tr>
<tr>
<td>All districts combined</td>
<td>117/290(40)</td>
<td>299/383(78)</td>
<td>416/673(62)</td>
</tr>
</tbody>
</table>

---

Table 2  Number (percentage) of concordant cases for nondate variables between cancer registry and hospital case notes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Diagnosis in 1983 (n = 117)</th>
<th>Diagnosis in 1988 (n = 299)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>116(99)</td>
<td>299(100)</td>
</tr>
<tr>
<td>District of residence</td>
<td>116(99)</td>
<td>299(100)</td>
</tr>
<tr>
<td>Tumour site</td>
<td>103(88)</td>
<td>263(88)</td>
</tr>
<tr>
<td>(colon or rectum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment details</td>
<td>100(86)</td>
<td>243(81)</td>
</tr>
<tr>
<td>All</td>
<td>93(80)</td>
<td>216(72)</td>
</tr>
</tbody>
</table>
During the course of our analysis we discovered that, owing to an error in the manual for registration officers, the registry's clinical staging system for colorectal cancer was incompatible with clinical staging and hence tumour stage could not be compared in this study.

Agreement on treatment data stood at 80% and 72% for 1983 and 1988 respectively. The registry recorded 107 cases as having had no treatment compared with only 58 cases recorded by the case notes. During the six month follow up the registry missed 24% of all surgical treatments and 37% of adjuvant therapy, radiotherapy, and chemotherapy (table 5).

**FACTORS ASSOCIATED WITH DISAGREEMENT**

**Source of registration** — Of the 385 disagreements over date of diagnosis 63(16%) were death certificate only registrations (table 6). Where case notes were the source of registration (84% of retrieved cases) most disagreements fell within 30 days of each other. Of the 49 cases with disagreement over tumour site, 16(33%) were death certificate only registrations. Death certificate only registrations accounted for 33(41%) of the 81 cases with disagreements over date of diagnosis of 30 days or more. Under-recording of treatment by the registry accounted for 87 disagreements with case notes (table 6); 38(44%) of these disagreements were death certificate only registrations whereas the remaining 49 cases were registered from case notes and of these, 18 had no treatment data recorded and 31 had partial treatment data recorded by the registry.

**Timing of registration** — None of the cases had a diagnosis to treatment time of greater than six months — that is, outside the registry follow up period.

**Survival** — Death certificate only registration seemed to be associated with missing treatment data in cases with poor survival (≤ one year) (table 7) but not with missing treatment data in cases surviving for longer. Overall, in 27(71%) of death certificate only registrations patient survival was more than three months, indicating that these cases could have been registered from case notes.

**Discussion**

**CASE NOTE RETRIEVAL**

Our retrieval rate varied strongly by district of residence, probably reflecting the differences in the operation of medical records departments among the districts. In district A case notes are retained for 10 years after the last date of attendance, except when the patient is deceased, when the retention period is foreshortened to six years. This may account for the poor retrieval rate for cases with a diagnosis in 1983. In district B case notes are kept for eight years after the last date of attendance, regardless of whether the patient is alive or not, before usually being destroyed. During the first two years they are stored at a single hospital. Thereafter they may be stored at another site. Case retrieval was highly efficient in this district; the slightly lower retrieval rate for 1983 may be the result of the destruction of case notes for cases with a diagnosis early in the year.

District C aims at transferring all case notes on to microfiche immediately after diagnosis. Although there is a policy of not destroying records, case note retrieval was poor. This is a very serious problem. In district D case notes are kept for five years after the last date of attendance, regardless of whether the patient is alive or not. They are then transferred to microfilm for indefinite retention. This takes about a month during which time notes can be sent for and accessed more or less immediately.

**AGREEMENT BETWEEN CANCER REGISTRY AND CASE NOTES**

Good agreement existed between the registry and hospital case note data for date of birth, date of death, sex, and district of residence. Disagreements on sex and district of residence arose as a consequence of one patient's...
Tables 6 and 7: Selected discorances between hospital case notes and cancer registry by source of registration. Figures are numbers (percentages)

<table>
<thead>
<tr>
<th>Source of registration</th>
<th>Case notes</th>
<th>Death certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing treatment data (n = 87):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No treatment data recorded</td>
<td>18(32)</td>
<td>38(68)</td>
</tr>
<tr>
<td>Partial treatment data recorded</td>
<td>31(100)</td>
<td>0(0)</td>
</tr>
<tr>
<td>Discordant ICD code (n = 49)</td>
<td>33(67)</td>
<td>16(32)</td>
</tr>
<tr>
<td>Discordant date of diagnosis (n = 385):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30 days</td>
<td>274(91)</td>
<td>30(9)</td>
</tr>
<tr>
<td>&gt; 30 days</td>
<td>48(59)</td>
<td>33(41)</td>
</tr>
<tr>
<td>All cases (n = 416)</td>
<td>350(84)</td>
<td>66(16)</td>
</tr>
</tbody>
</table>

The Thames cancer registry number having been reassigned to another patient. The accuracy of recording of date of death rose from 86% to 93% between 1983 and 1988.

The major disagreements between the cancer registry and case notes arose over treatment, tumour site, and date of diagnosis. Each of these has serious implications for the reliability of cancer registry statistics. Disagreements over date of diagnosis will bias survival data, disagreements over site will affect incidence data and trends, and disagreements over treatment undermine the case for using registry data to evaluate care.

This study looked at those aspects of the registration process which could potentially contribute to disagreements between case note data and cancer registry data. The failure to observe registry policies on case ascertainment, active follow up, and coding criteria were all shown to contribute, and these are discussed below.

**Disagreements arising from the primary data source**

The major cause of disagreement between the case note data and the registry data was that the registry's use of death certificates as the primary data source for registration. This finding highlights a previously unreported problem of using case notes as the standard in reliability studies. Our success in obtaining case notes for 66 of the 150 death certificates only registrations in the original sample indicates that the registry is failing to ascertain case notes on death certificate initiatedregistrations. This finding is consistent with the dramatic rise in the proportion of death certificate only registrations in the South Thames regional health authorities that took place during 1982–8, from 5% to 30% of registrations (A M Pollock, N Vickers, personal communication).

However, disagreement between case note data and the registry data on date of diagnosis, tumour site, and treatment could be only partly attributed to the registry having used death certificates as their source of registration. Of the 10% of registrations assigned to the wrong year of diagnosis, only half were death certificate only registrations; of the 49 cases with discordant data on tumour site, 16(33%) were death certificate only registrations; and of the remaining differences were due to a combination of coding error, data omission, and failure to observe the registry policy of six month active follow up of case notes.

**Disagreements arising from differences in coding practice**

Absolute agreement for date of diagnosis was very low (12% for both 1983 and 1988). This contrasts with the study by Gulliford et al of registration of bladder cancer, which, once death certificate only cases had been excluded, showed good agreement between case notes and the Thames cancer registry for date of diagnosis. Although we followed the Thames cancer registry's criteria for date of diagnosis, it became apparent that not all coding clerks had done so, and we therefore considered estimates within 30 days of one another as being in broad agreement. This resulted in rates of agreement over date of diagnosis rising to 83% and 77% respectively in 1983 and 1988.

**Differences arising from failure of follow up**

The Thames cancer registry under-recorded 20% of all surgical treatments and 37% of all adjuvant therapy, higher than reported by Gulliford et al, who quoted agreement rates between case notes and the registry of 15% for chemotherapy. No treatment was administered after the registry's follow up period, indicating that in those cases registered by case notes there was a failure of the registry's six month follow up policy.

**Differences arising from data omission**

Most (38/56, 68%) cases with all treatment data missing were recorded by death certificate only; 71(27/38) of these showed the patient's survival longer than three months, indicating that the patient might have been registered in life, had more efficient registration procedures been in place. Among cases with missing treatment data in patients surviving more than a year, 61(23/38) were registered from case notes. Further research is warranted to investigate these omissions from the notes.

**Outcome of study**

During this study we had regular meetings with senior registry staff to feed back our results and to learn about the impact of the registry's own quality assurance programmes on registration. The anomaly between the Thames cancer registry's staging system and the Dukes' classification meant that we were unable to evaluate the effectiveness of health care for this period. As a result of our work the anomaly has been rectified. Our findings were presented to the local audit coordinators and clinicians at all four district hospital sites either at combined

Table 7: Survival times of patients with treatment data omitted from the cancer registry by source of registration. Figures are numbers (percentages)

<table>
<thead>
<tr>
<th>Survival time (days)</th>
<th>Source of registration</th>
<th>Case notes</th>
<th>Death certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 30</td>
<td>1(11)</td>
<td>8(89)</td>
<td></td>
</tr>
<tr>
<td>&gt; 30 days – &lt; 91 days</td>
<td>1(25)</td>
<td>3(75)</td>
<td></td>
</tr>
<tr>
<td>&gt; 90 days – 1 year</td>
<td>2(16)</td>
<td>12(86)</td>
<td></td>
</tr>
<tr>
<td>&gt; 1 year</td>
<td>23(61)</td>
<td>15(40)</td>
<td></td>
</tr>
</tbody>
</table>
Reliability of Thames cancer registry data on an colorectal cancer

clinical and pathology meetings or at post-graduate meetings. Purchasers attended some of these meetings. We recently completed two major studies of death certificate only registrations in south east England,20 21 and the Thames cancer registry is undertaking a major project to reduce the number of death certificate only registrations. It is hoped that this project will enhance levels of agreement between the Thames cancer registry and hospital case notes.

This study shows how cancer registries might use the process of registration to check data accuracy and completeness. We showed how disagreements between registry data and case notes over date of diagnosis, survival, tumour site, and treatment reflect a failure of registry policy in three areas: retrospective follow up of death certificate initiated registrations; six month active follow up of cases after diagnosis of cancer, and registry coding criteria for date of diagnosis. Registries need to ensure that they have good quality monitoring programmes of their data collection procedures if data quality is to be assured for epidemiological surveys and health services research.

Reliability of data of the Thames cancer registry on 673 cases of colorectal cancer: effect of the registration process.

A M Pollock and N Vickers

*Qual Health Care* 1995 4: 184-189
doi: 10.1136/qshc.4.3.184

Updated information and services can be found at:
http://qualitysafety.bmj.com/content/4/3/184

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/