THE NEW ZEALAND MEDICAL JOURNAL
Journal of the New Zealand Medical Association

An audit of colon cancer data on the New Zealand Cancer Registry

Ruth Cunningham, Diana Sarfati, Sarah Hill, Diane Kenwright

Abstract

Aims This study aims to assess the reliability of New Zealand Cancer Registry data on colon cancer.

Methods Data from a review of the clinical records of 642 people diagnosed with colon cancer between 1996 and 2003 were used to audit the data held on these individuals by the New Zealand Cancer Registry (NZCR).The record review data were treated as the "gold standard".

Results Age at diagnosis (measured in years) recorded by NZCR was 96% accurate, and date of diagnosis was within 6 weeks of the clinical date of diagnosis in more than 97% of cases. Overall tumour site was recorded with more than 95% accuracy, with 86% accuracy for tumour sub-site within the colon. Tumour grade was only recorded consistently by the NZCR from 1999 onwards, from which time the NZCR was 83% accurate for tumour grade. Tumour stage was the least accurate variable studied, with 80% accuracy. The NZCR data quality improved over the period of this study.

Conclusions The accuracy of the NZCR appears to be similar to that found in comparable audits of cancer registries, with stage being the hardest variable for registries to collect accurate information on. NZCR data could be improved by improving the quality of information provided to the registry.

Cancer registries have a vital role in cancer control. The New Zealand Cancer Control Strategy[1] has as one of its six goals improving the effectiveness of cancer control through research and surveillance, recognising the central role of information in cancer control. The New Zealand Cancer Registry (NZCR) is a population-based register of all primary malignancies diagnosed in New Zealand (excluding basal and squamous cell skin cancers), and is the primary source of information on cancer incidence in New Zealand.

The International Agency for Research on Cancer (IARC) has identified five main areas of quality to be considered in assessing cancer registries: completeness of cover, completeness of detail, accuracy of detail, accuracy of reporting, and accuracy of interpretation.[2]

Two studies in New Zealand have assessed the completeness of coverage of the NZCR. Dockerty et al[3] looked at the accuracy and completeness of child cancer registrations between 1990 and 1993 using data from the Children's Cancer Registry and hospital admissions and discharges, and found that the NZCR ascertained 97% of cases of childhood cancer over this period, but nearly 10% of cases reported by the NZCR were not in fact confirmed as incident cases of childhood cancer.

A recent audit of lung cancer treatment in Auckland and Northland used regional clinical databases to find additional cases of lung cancer beyond those known to the NZCR, and found that 66 out of 565 cases meeting the eligibility criteria for the study were not known to the NZCR.[4]

Internationally, reviews of medical records have been used as a gold standard against which to audit cancer registry data accuracy,[5–11] although it is also possible to use clinical databases, such as were used by Stevens et al to review the accuracy of lung cancer data.[4]

The objective of this audit was to compare the data accuracy of the NZCR against data extracted from the clinical records of 642 people registered with colon cancer between 1 January 1996 and 31 December 2003 as part of a study of Māori /non-Māori colon cancer survival differences. We did not assess the completeness of cover of the NZCR colon cancer data as no suitable dataset for comparison could be readily accessed; neither did we examine completeness of registry detail as this is more appropriately assessed internally by cancer registries.

Approval for this study was granted by the New Zealand Multi-Region Ethics Committee.

Methods

The New Zealand Cancer Registry is a population-based register of all primary malignancies diagnosed in New Zealand. New cancer diagnoses are reported to the Registry mainly by laboratories, which are required by the Cancer Registry Act 1993 to send copies of pathology reports diagnosing cancer. A small proportion of cancer registrations are derived from hospital discharge reports (public and private), death certificates, and coroners' reports.

The NZCR system was upgraded in 2001, with the new database going live in December 2001, at which time data from 1999 and 2000 were still being coded. Data from 1999 onwards are coded using the updated system, which has an increased number of fields and more complete details of cancer stage and morphology (personal communication, S Hanna, 2007). NZCR data are entered by trained coders who specialise in particular cancer sites.

This audit used detailed clinical record data collected for a separate study on colon cancer survival. For the record review, incident cases of colon cancer were identified from the New Zealand Cancer Registry. All eligible Māori cases and a randomly sampled equal number of non-Māori cases were included in the study sample.

The final sample meeting the eligibility criteria (see Figure 1) and with data available included 308 Māori and 334 non-Māori New Zealanders with colon cancer.

Pathology reports were obtained and reviewed for all study patients. Clinical data were obtained from both public hospital and private specialists' records by one of the authors (SH). Data extraction was carried out according to standardised criteria. All data were double-entered and discrepancies were checked.

NZCR data were checked against clinical record data for the same individuals. Clinical records were regarded as a "gold standard"—that is they were regarded as the definitive benchmark against which registry data could be checked for accuracy. Data were analysed using Stata version 10.[12]

Discrepancies between the NZCR and hospital record were examined for the following variables: age at diagnosis, sex, date of diagnosis, tumour site within the body, site within the colon, and tumour grade and stage.

The percentage of discrepancies between the record and NZCR datasets are presented as estimates for the total population with colon cancer in order to give an indication of the overall accuracy of the data on colon cancer held by the NZCR, and for the Māori and non-Māori samples combined (raw data).

Population estimates were generated by weighting the data for Māori and non-Māori samples according to the proportion of total colon cancer notifications represented by each ethnic group between 1996 and 2003.

Figure 1. Eligibility criteria

Study Eligibility Criteria

- newly diagnosed cancer of the colon registered between 1 January 1996 and 31 December 2003
- primary tumour site in the colon: ICD10-AM site codes C18-C19, not including C18.1 (appendix)
- no previous diagnosis of colon cancer
- morphology consistent with or specific to adenocarcinoma: (ICD-O morphology codes 8000, 8010, 8020, 8021, 8050, 8140, 8144, 8145, 8210, 8211, 8260, 8261, 8262, 8263, 8470, 8471, 8472, 8473, 8480, 8481, and 8490)
- aged 25 years or over at diagnosis
- usually resident in New Zealand
- diagnosis made prior to death

In some cases, variables were assigned differently by NZCR and in record review. Table 1 sets out the different methods used to assign each variable, and any methods used in this study to facilitate comparison of the two datasets.

Table 1. Differences in methods of assigning variables between datasets

Variable	NZCR	Clinical record	Method for dealing with differences
Age at diagnosis	See date of diagnosis. Date of birth recorded under NHI	See date of diagnosis. Date of birth recorded under NHI	nil
Sex	As recorded under NHI	As recorded under NHI	nil
Date of diagnosis	Date of pathology report, or date of hospital admission, or date of death (if post-mortem diagnosis)	Clinical date of diagnosis (date cancer confirmed)	Accuracy to within 6 weeks reported (exact correlation not expected because of different definitions)
Tumour site within body	Based on pathology report or hospital discharge data	Based on all available information including investigation results, examination findings and surgical and pathology reports	nil
Tumour site in colon	Based on pathology report (pathologist relies on clinical information provided by the requesting clinician). Categories based on ICD-9 and ICD-10AM	Based on all available information including investigation results, examination findings and surgical and pathology reports. Categories based on ICD-	Categories overlapping (used by NZCR) and synchronous (used in record review) excluded from analysis as not comparable

		10AM, but also including synchronous category	
Tumour grade	Based on pathology report, only recorded reliably from 1999 (prior to this recorded in free-text field), recorded as unknown if no grade stated on report, least differentiated grade recorded	Based on pathology report, recorded as moderate if no grade stated* (47 cases), least differentiated grade recorded	Those registered prior to 1999 and those with no grade stated on the report were excluded from comparisons
Tumour stage/ extent of disease	SEER summary staging system based on pathology report (principally) and any investigations within four months of diagnosis (investigation reports not reliably received by NZCR), assigned according to the furthest extent of known involvement, reported as 'extent of disease'	TNM staging (pathological) based on all available clinical data, including all investigations within four months of diagnosis, assigned according to the furthest extent of known involvement	Record data converted to SEER summary staging (NB required recoding from pathology reports as the systems do not map one to one)

*Based on pathologist's advice that pathologists may only make note of exceptional grade (well and poorl;y differentiated cancers) in their reporting.

Results

Table 2 shows the discrepancies between the NZCR and record data for each of the six fields examined. Table 3 shows the trends over time in the quality of cancer registry data for the main variables examined.

There was good agreement between the datasets with respect to age at diagnosis (measured in years), sex, and date of diagnosis (within 6 weeks). Agreement was less complete for the other variables examined.

Table 2. Discrepancies between cancer registry and clinical record

Field	Population estimate (weighted)^ % discrepancies	No. records with data available (raw data)	No. discrepancies
Age at diagnosis (years)	3.6%	642	20
Sex	0.6%	642	2
Diagnosis date (exact)*	71.2%	634	436
Diagnosis date within 6 weeks*	2.7%	634	15
Tumour site within body**	4.6%	776	51
Tumour site within colon***	13.6%	600	78
Tumour grade (1999 onwards)****	17.4%	420	78
Extent of disease (stage)	19.7%	642	122

*Excluding those with derived diagnosis date in record data (n=8); **Includes all cancers in original sample reported by the NZCR to be colon cancer, of which 51 were not in fact confirmed colon cancer primaries. A further 83 of the 776 were excluded from the final study sample for other reasons; ***Excluding synchronous and overlapping categories (n=42); ****Excluding grade not stated on pathology report, no field for tumour grade prior to 1999; ^Weight=proportion of ethnic group in total NZCR colon cancer population. Māori weight=0.0256; Non-Māori weight=0.9744.

Table 3. Trends over time in discrepancies

Field	Population estimate (weighted)^ % discrepancies	No. records with data available (raw data)	No. discrepancies
Tumour site within body ** (1996-1998)	6.3%	274	21
Tumour site within body (1999-2001)	3.3%	302	16
Tumour site within body (2002-2003)	3.8%	200	15
Tumour site in colon*** (1996-1998)	15.0%	222	26
Tumour site in colon*** (1999-2001)	11.4%	152	30
Tumour site in colon*** (2002-2003)	10.1%	168	22
Tumour grade**** (1999-2001)	11.2%	238	29
Tumour grade**** (2002-2003)	19.1%	162	29
Extent of disease (1996-1998)	26.6%	222	61
Extent of disease (1999-2001)	17.3%	252	38
Extent of disease (2002-2003)	11.5%	168	23

**Includes all those reported by the NZCR to be colon cancer; ***Excluding synchronous and overlapping categories; ****Excluding grade not stated on report; ^Weight=proportion of ethnic group in total NZCR colon cancer population. Māori weight=0.0256; Non-Māori weight=0.9744.

Tumour site—Of the 776 patients whose records were examined (all reported by the NZCR to have colon cancer), 51 (7%) did not in fact have a diagnosis of colon cancer. In 19 cases, the primary tumour was not located, in 29 cases the tumour was in the rectum rather than the colon, and there was one case in each of the oesophagus, small bowel, and stomach. A further 83 cases were not eligible for the study for other reasons (principally lack of histological diagnosis), resulting in a final sample of 642.

Tumour site within colon—There was a discrepancy in tumour site within the colon between the two datasets in approximately 13% of cases (once the categories synchronous and overlapping had been excluded), and this remained similar over the study period. Table 4 shows the actual discrepancies found between the datasets. Most of the miscoding of tumour site is miscoding to an adjacent site (between right and left colon, between left colon and rectosigmoid junction).

Tumour grade—Following the upgrade of the NZCR in 2001 (affecting data from 1999), grade information was 83% accurate. Table 5 shows the actual discrepancies between the datasets after exclusion of those diagnosed prior to 1999. The most common source of discrepancy was where NZCR assigned an unknown grade, while the record review identified a grade.

In other cases, different grades were reported by the NZCR and record data extraction. This most often occurred when the pathology report noted more than one grade in different parts of the tumour, in which case the record review recorded the higher grade (less differentiated) while the NZCR often recorded the lower grade.

Table 4. Comparison of record and registry data for cancer site*

Tumour site: cancer registry	Tumour site: clinical record					
	R colon	L colon	Rectosigmoid	Synchronous	unknown	Total
R colon**	**248**	11	0	12	0	**271**
L colon	12	**206**	38	15	0	**271**
Rectosigmoid	0	10	**62**	3	0	**75**
Overlapping***	4	2	1	5	0	**12**
unknown	1	5	1	0	**6**	**13**
Total	**265**	**234**	**102**	**35**	**6**	**642**

*Bold numbers indicate the number of records with agreement between the two datasets, while the other numbers indicate records with discrepancies; **R(ight) colon: caecum, ascending colon, hepatic flexure, transverse colon; L(eft) colon: splenic flexure, descending colon, sigmoid colon; ***This category was not used in record review.;

Table 5. Comparison of record and registry data for tumour grade (excluding no grade stated, and prior to 1999)

Tumour grade: cancer registry	Tumour grade: clinical record				
	Well differentiated	Moderately differentiated	Poorly differentiated	Unknown	Total
Well differentiated	**32**	3	2	0	**37**
Moderately differentiated	3	**243**	12	0	**258**
Poorly differentiated	0	3	**64**	0	**67**
Undifferentiated*	0	0	1	0	**1**
unknown	4	43	7	**3**	**57**
Total	**39**	**292**	**86**	**3**	**420**

*This category was not used in record review.

Table 6. Comparison of record and registry data for extent of disease 1996–1998

Extent of disease: cancer registry	Extent of disease: clinical record				
	1 Localised	2 Regional spread*	3 Metastatic spread	5 Unknown	Total
1 Localised	**37**	14	0	0	**51**
2 Regional spread*	11	**88**	15	0	**114**
3 Metastatic spread	0	7	**35**	0	**42**
5 unknown	6	4	4	**1**	**15**
Total	**54**	**113**	**54**	**1**	**222**

*Direct extension or lymph node involvement.

Tumour stage (extent of disease)—The proportion of discrepancies between the two datasets for tumour extent of disease was approximately 20% and this reduced over the time period of the study. Tables 6 and 7 show that there were two main areas of discrepancies: between localised and regional disease (with discrepancies in both directions) and between regional and advanced disease (with the NZCR showing a less advanced extent than the clinical record).

Table 7. Comparison of record and registry data for extent of disease 1999–2003

Extent of disease: registry	Extent of disease: clinical record					
	B Localised	C Direct regional spread	D Lymph node involvement	E Metastatic spread	F Unknown	Total
B Localised	**83**	11	1	1	0	**96**
C Direct regional spread	11	**58**	0	0	0	**69**
D Lymph node involvement	1	2	**125**	14	0	**142**
E Metastatic spread	0	4	6	**86**	0	**96**
F unknown	3	0	2	5	**7**	**17**
Total	**98**	**75**	**134**	**106**	**7**	**420**

Discussion

New Zealand Cancer Registry data is used for wide a range of applications, including research, policymaking, and health service planning. Any such work depends for its accuracy on the quality of NZCR data, and so it is important to know that data provided by the NZCR are reliable in terms of demographic and diagnostic details.

This study found that cancer registrations for colon cancer on the NZCR were highly accurate with respect to demographic details, but less so for details relating the site, grade and stage of the tumour. The accuracy of the NZCR appears to be similar to that found in comparable audits of cancer registries in the United States, the United Kingdom, and the Netherlands,[5–11] with stage being the hardest variable for registries to collect accurately.

The recent audit of lung cancer registrations in New Zealand conducted by Stevens et al also found that stage was the least accurate variable.[4] Encouragingly, NZCR data improved following changes to the registry in 2001.

The key limitation of this study is that it assumes that perfect information is captured by the clinical record, which is of course not the case. Identifying clinical and pathological inaccuracies recorded in clinical records is beyond the scope of this study. However, such errors will be largely consistently recorded by both record review and the NZCR because data for the NZCR are usually drawn either directly from clinical sources (e.g. the pathology report) or from data sources which are extracted from clinical record data (e.g. hospital discharge data).

It is also assumed that data extracted from clinical records are recorded without errors. All data were extracted by the same individual according to pre-designated rules, and were double entered to avoid data-entry errors. Original pathology reports were also referred to in order that particular discrepancies could be better understood. Both of these methods help to limit the possibility of errors arising in the clinical record data.

Some discrepancies were to be expected given the different ways in which variables were defined in the two datasets. This was particularly the case for date of diagnosis, where completely different definitions were used, and this difference also affected the comparability of age at diagnosis.

Other discrepancies found relate to the nature of information available to the NZCR. NZCR coding of tumour site (and most other information) relies almost solely on pathology reports. Pathologists rely in turn on surgeons for an indication of tumour

site to be noted with the specimen, as it can be difficult to accurately site a segment of resected colon. For this reason pathology reports may not give an accurate reflection of tumour site. Data drawn from clinical records included information from operation notes and scan reports, providing a better characterisation of tumour site. The different sources of data may explain most of the discrepancies seen. The best way to overcome this problem is through improving the quality of reporting to the NZCR.

The Australasian College of Pathologists is currently developing a *pro forma* for synoptic reporting of colorectal cancer specimens (a standardised template for pathology reporting), which if adopted will provide more consistent information for NZCR coders. A *pro forma* for breast cancer specimen reporting has been used since the introduction of breast cancer screening in New Zealand. Communication with surgeons about the need for adequate clinical details on pathology request forms may also help to improve the quality of information available to the NZCR.

Further discrepancies relate to the systems used by the NZCR, particularly in relation to staging. The SEER summary staging system[13] is used by the NZCR. In contrast, the TNM system[14] is frequently used by pathologists and other clinicians. TNM is a clinical staging system which categorises cancer spread according to three characteristics: the primary tumour (T), involvement of regional lymph nodes (N) and presence or absence of distant metastases (M). The TNM and summary stage systems are not directly comparable.

As can be seen from Table 8, the distinction between localised and regional disease in the SEER system divides the T3N0M0 (IIa) category in two for colon cancer, with some cancers in this category counting as localised and some as regionally advanced. Problems with the use of summary staging systems have been noted by other authors,[8] and confusion over the application of the local/regional distinction for colon cancer in the SEER system is reported by NZCR coders (personal communication, C Bainbridge, 2007), which may explain the discrepancies between localised and regional disease found between the clinical record and NZCR datasets.

The TNM and SEER summary staging systems have different strengths. The SEER system is specifically designed for use by cancer registry coders. It has the advantage of being less complex than other staging systems and relatively stable over time. The TNM system on the other hand is a dynamic system designed for use by clinicians. TNM stage is assigned based on clinical and pathological observations, and is intended to give good prognostic information to clinicians.

Table 8. SEER summary stage and equivalent TNM stage for colon cancer*

SEER summary stage	Description	Equivalent TNM stage
Localised	Invasive tumour confined to colon. Includes tumour extension through muscularis propria and subserosal tissue, but not serosal surface.	Stage I and IIa: T1–T3 N0 M0
Regional	Tumour extension outside colon and/or invasion of regional lymph nodes. Includes local tumour extension into serosal surface, pericolic or mesenteric fat **	Stage IIa and IIb and III: T3–T4 / Any N Any T / N1,2 M0
Distant	Tumour spread to distant organs or lymph nodes.	Stage IV: Any T Any N M1

*From SEER Program Coding and Staging Manual 2007[15]; **Also adjacent tissues/connective tissue/fat, mesentery, mesocolon, retroperitoneal fat, gastrocolic ligament, greater omentum and any other abdominal or pelvic organs.

The NZCR is reliant on pathologists recording TNM stage on pathology reports if it is to collect information on TNM stage. However the problems with assigning SEER summary stage in colon cancer, and the problems for clinicians in interpreting data which use the SEER system, mean that the TNM stage should also be recorded by the NZCR whenever possible.

The UK guidelines for colorectal cancer management recommend that TNM stage is always recorded on colorectal pathology reports.[16] If such a requirement was adopted in New Zealand it would be much easier for the NZCR to reliably collect this information. In the United States a collaborative staging system is being developed specifically to overcome problems of incompatibility between staging systems,[17] and the possibility of using this system in New Zealand could be explored.

Cancer registry data is the main source of information on the incidence of one of New Zealand's major causes of death and disability. It is therefore very important that the quality of the data provided by the NZCR is monitored and improved.

New Zealand is currently in the process of developing extensions and improvements to cancer data collections, which will provide an opportunity to address some of the issues raised here around the flow of information from clinicians to the registry, as well as ensuring the appropriate systems for collecting information such as stage are used. The recently announced intention to introduce colorectal cancer screening is also likely to provide an impetus to improve colorectal cancer data collection.

Competing interests: None known.

Author information: Ruth Cunningham, Public Health Medicine Registrar; Diana Sarfati, Senior research fellow, Department of Public Health, University of Otago, Wellington; Sarah Hill, Senior Research Fellow, Department of Public Health, University of Otago, Wellington; Diane Kenwright, Pathologist, Department of Anatomic Pathology, Wellington Hospital, Wellington

Acknowledgements: The authors thank Susan Hanna and Christine Bainbridge from NZHIS for their helpful advice on the workings of the NZCR, Emma Britton for assistance in obtaining pathology reports and checking data, and the members of the team involved in designing and carrying out the Ethnic Disparities in Colon Cancer Survival study from which data were drawn for this paper.

The members of the study team are Diana Sarfati, Sarah Hill, Gordon Purdie, Bridget Robson, Donna Cormack, Tony Blakely, Elizabeth Dennett, and Kevin Dew. The authors also acknowledge the Cancer Society of New Zealand for providing funding for the above study (grant 05/16).

The Ministry of Health provided funding for a report on New Zealand Cancer Registry data quality on which this paper is based (the views expressed in this paper do not necessarily reflect those of the Ministry of Health). The Australasian Faculty of Public Health Medicine training programme provided a salary subsidy for Ruth Cunningham.

Correspondence: Diana Sarfati, Department of Public Health, University of Otago, Wellington, PO Box 7343, Wellington South, New Zealand. Fax: +64 (0)4 3895319; email diana.sarfati@otago.ac.nz

References:

1. Minister of Health. The New Zealand Cancer Control Strategy. Wellington: Ministry of Health and the New Zealand Cancer Control Trust; 2003.
2. Skeet RG. Quality and quality control. In: Jensen OM, Parkin DM, editors. Cancer Registration: Principles and Methods. Lyon: International Agency for Research on Cancer, 1991:101–7.
3. Dockerty JD, Becroft DM, Lewis ME, Williams SM. The accuracy and completeness of childhood cancer registration in New Zealand. Cancer Causes & Control. 1997;8:857–64.
4. Stevens W, Stevens G, Kolbe J, Cox B. Comparison of New Zealand Cancer Registry data with an independent lung cancer audit. NZMJ 2008;121(1276). http://www.nzmj.com/journal/121-1276/3119
5. Liu W-L, Kasl S, Flannery JT, et al. The accuracy of prostate cancer staging in a population-based tumor registry and its impact on the Black-White stage difference (Connecticut, United States). Cancer Causes and Control. 1995;6:425–30.
6. Thoburn KK, German RR, Lewis M, et al. Case completeness and data accuracy in the Centres for Disease Control and Prevention's National Program of Cancer Registries. Cancer. 2007;109:1607–16.
7. Brewster DH, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. European Journal of Cancer. 2002;38:414–7.
8. Gulliford M, Bell J, Bourne H, Petruckevitch A. The reliability of Cancer Registry records. British Journal of Cancer. 1993;67:819–21.
9. Middleton RJ, Gavin AT, Reid JS, O'Reilly D. Accuracy of hospital discharge data for cancer registration and epidemiological research in Northern Ireland. Cancer Causes & Control. 2000;11:899–905.
10. Lapham R, Waugh NR. An audit of the quality of cancer registration data. British Journal of Cancer. 1992;66:552–4.
11. Schouten LJ, Langendijk JA, Jager JJ, van den Brandt PA. Validity of the stage of lung cancer in records of the Maastricht cancer registry, The Netherlands. Lung Cancer. 1997;17:115–22.
12. StataCorp. Stata statistical software: release 10. College Station, Texas: StataCorp, 2007.
13. Young J, Roffers F, Gloeckler Ries L, Fritz A, editors. SEER Summary Staging Manual - 2000: Codes and Coding Instructions. Bethesda, MD: National Cancer Institute; 2000.

14. AJCC. American Joint Committee on Cancer: Cancer Staging Manual. Sixth Edtion ed. New York: Springer-Verlag; 2002.
15. Johnson C, Adamo M, editors. SEER Programme Coding and Staging Manual. Bethesda, MD: National Cancer Institute; 2007.
16. ACPGBI. Guidelines for the Management of Colorectal Cancer, Third Edition. London: The Association of Coloproctology of Great Britain and Northern Ireland; 2007.
17. Collaborative Staging Task Force of the American Joint Committee on Cancer. Collaborative Staging Manual and Coding Instructions. version 01.03.00 Jointly published by American Joint Committee on Cancer (Chicago, IL) and U.S. Department of Health and Human Services (Bethesda, MD); 2006.