Same Question, Different Data Source, Different Answers? Data Source Agreement for Surgical Procedures on Women with Breast Cancer

Même question, sources de données différentes, réponses différentes? Concordance entre les sources des données pour les interventions chirurgicales pratiquées sur les femmes atteintes de cancer du sein

by D. TURNER, PHD
Department of Epidemiology and Cancer Registry
CancerCare Manitoba
Winnipeg, MB

K.J. HILDEBRAND, MSC
Department of Epidemiology and Cancer Registry
CancerCare Manitoba
Winnipeg, MB

K. FRADETTE
Department of Epidemiology and Cancer Registry
CancerCare Manitoba
Winnipeg, MB

S. LATOSINSKY, MD, MSC, FRCSC
Division of Surgical Oncology
Health Sciences Centre
Winnipeg, MB

Abstract

This study assessed the accuracy of the Manitoba Cancer Registry (MCR) and two administrative data sources, the Manitoba Health hospital discharge file and the Manitoba Health medical claims file, for capturing surgical procedures related to the treatment of breast cancer. The study cohort included all women diagnosed in Manitoba with invasive or in situ breast cancer between 1995 and 1999. The surgical procedures of interest were mastectomy, breast conserving surgery and axillary node dissection. Analysis focused on assessing concordance between data sources following record linkage. Agreement was measured using the kappa statistic, and chart reviews of discordant information were completed to identify the more reliable data source and to validate data files. The effect of using each data set alone to calculate procedure rates was determined to identify any clinically important differences arising from the choice of data source. Results indicate that capture of breast cancer patients using administrative data sets alone can be quite good and that the population-based cancer registry is superior to other administrative data sets for capturing surgical treatment information on cancer cases.

Résumé

Cette étude visait à évaluer l'exactitude du Manitoba Cancer Registry (MCR) et de deux sources de données administratives – le dossier des congés des hôpitaux et le dossier des demandes d'indemnisation de Santé Manitoba – pour le recensement des interventions chirurgicales relatives au traitement du cancer du sein. La cohorte de l'étude comprenait toutes les femmes chez qui on a diagnostiqué un cancer du sein invasif ou in situ au Manitoba entre 1995 et 1999. Les interventions chirurgicales concernées étaient la mastectomie, la chirurgie mammaire conservatrice et l'évidement ganglionnaire axillaire. L'analyse visait surtout à évaluer la concordance entre les sources de données après avoir établi des liens entre les dossiers. La concordance a été mesurée en utilisant l'analyse statistique kappa et en examinant les dossiers médicaux pour les données discordantes en vue de déterminer la source de données la plus fiable et de valider les fichiers de données. L'utilisation isolée de chaque ensemble de données pour calculer les taux d'intervention chirurgicale a permis de repérer les différences cliniques significatives découlant du choix de la source de données. Les résultats indiquent que la saisie de données sur les patientes atteintes de cancer du sein en uti-

lisant les ensembles de données administratives séparément peut s'avérer très bonne et que le registre du cancer axé sur la population est supérieur aux autres ensembles de données administratives pour ce qui est de la saisie de renseignements sur les traitements chirurgicaux liés au cancer du sein.

Introduction

A number of data sources have been employed in the many published studies of breast cancer diagnosis and surgery, including prospectively collected clinical data sets, retrospective chart review, administrative data and cancer registries (Malin et al. 2002a). Good-quality clinical data sets are not widely available, however, and often cover only small populations or even a single hospital, while chart review or abstraction is resource intensive and costly. Consequently, administrative data sets are employed extensively because of their availability, coverage and low cost, but their accuracy has been questioned (Pinfold et al. 2000). Administrative data errors may result from incomplete information available to the coder, transcription errors during data capture or incorrect coding due to differences in the interpretation of coding rules (Middleton et al. 2000); indeed, all data sources require some form of quality control.

This study assessed the accuracy of the Manitoba Cancer Registry (MCR) and two administrative data sources, the Manitoba Health hospital discharge file and the Manitoba Health medical claims file, for capturing surgical procedures related to the treatment of female breast cancer.

Methods

The study cohort, identified from the population-based MCR, comprised all women diagnosed in Manitoba with invasive or in situ breast cancer between 1995 and 1999. Recent case ascertainment studies supported jointly by Statistics Canada and CancerCare Manitoba indicate that the MCR captures more than 99.5% of all cancers and 100% of breast cancers in the province. For women with multiple tumours, one "index" tumour was chosen using the following hierarchy: earliest diagnosis, highest stage and largest size. If these criteria were identical, then the index tumour was randomly selected.

Treatment information is routinely collected for each primary tumour in the MCR and was recorded according to ICD9-CM coding standards for the study period. MCR coders are certified as either health records technicians or health information technologists and receive one year of intensive on-the-job training in oncology

coding. The hospital discharge file includes records of all inpatient and day surgery admissions to Manitoba's acute and chronic care hospitals. All treatments were coded by hospital coders trained as health records technicians or health information technologists using ICD9-CM standards. The medical claims file is generated by fee-for-service claims made by Manitoba physicians. Staff in physicians' offices and claims-processing centres focus on coding jurisdiction-specific fee codes (tariff codes) for medical activities following rules specified by Manitoba Health.

The ability of the data sets to accurately capture mastectomy, breast-conserving surgery (BCS, also known as lumpectomy) and axillary node dissection (AND) was investigated. Relevant codes for each data set are shown in Table 1. BCS was defined in two ways: (1) by ICD9-CM codes 85.21–85.23, as suggested by others (Iscoe et al. 1997; C. DeCoster, Community Health Sciences, University of Manitoba, personal communication 2005); and (2) by ICD9-CM codes 85.21–85.23 and 85.12. Tariff codes for BCS were not introduced until 1999, and thus could not be captured from the medical claims data for the study period.

The ability of the data sets to accurately capture mastectomy, breast-conserving surgery (BCS, also known as lumpectomy) and axillary node dissection (AND) was investigated.

Procedures associated with the cohort that occurred within one month prior to one year after diagnosis were extracted from each data source. For multiple procedures, the most extensive procedure within one year of diagnosis was considered definitive. For example, if a mastectomy followed a BCS, the mastectomy was selected.

Analysis focused on assessing concordance between data sources following record linkage. Agreement was measured using the kappa statistic, which determines non-random agreement between two measurements of a categorical variable. Agreement indicated by kappa coefficients <0.00 is considered poor; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect (Landis and Koch 1977). Chart reviews of discordant information were completed to identify the more reliable data source and to validate data files. The effect of using each data set alone to calculate procedure rates (the total number of procedure occurrences divided by the total number of women in the defined subgroup from the original cohort) was determined to identify any clinically important differences arising from the choice of data source.

TABLE 1. Breast surgery procedures defined by codes

Procedure	ICD9-CM code*	Tariff code*
BCS, Definition 1[1]	85.12, 85.22, 85.23	0442
BCS, Definition 2[2]	85.12, 85.21, 85.22, 85.23	0442
Axillary node dissection (AND) ("Regional node dissection" in MCR)	40.3, 40.51	2658
Breast conservation surgery + AND		0443
Simple mastectomy (removal of breast only, not nodes)	85.41, 85.42	0449, 0457, 0477, 0478
Modified radical mastectomy (simple mastectomy + AND)	85.43, 85.44	0471
Radical mastectomy (includes removal of chest wall – pectoralis major muscle)	85.45–85.48	0470

* Code descriptors found in Appendix A online at http://www.longwoods.com/product.php?productid=19140&cat=499&page=1

[1] ICES Definition

[2] Current Study Definition

Results

The MCR captured information on 4,079 cases of breast cancer diagnosed in Manitoba in 3,956 women between 1995 and 1999. Of these women, 3,950 (99.8%) had a valid Personal Health Identification Number (PHIN), the key variable used in record linkage.

A surgical treatment record was found in the hospital discharge file for 95.7% of the women in the cohort, where only 33 (<1%) did not have an ICD9-CM breast cancer diagnostic code. Similarly, a medical claims record indicating breast cancer surgery was found for 96.2% of the women in our cohort, and only 22 (<1%) did not have a breast cancer diagnosis coded in the claim record. Agreement between these databases, in terms of their ability to capture breast cancer surgery, is shown in Table 2. All kappas indicated substantial or almost perfect agreement between data sets.

A review was conducted of 60 charts from the 345 patients recorded as having mastectomy in the medical claims database but not in the MCR. All but two of these patients were confirmed to have had BCS as their surgical procedure. A chart review of the discordant MCR and medical claims file AND cases found that the MCR always reflected what was described in the operative report. The majority of the discordance between the MCR and the hospital data was therefore attributed to an AND being performed but the hospital discharge file failing to record it.

TABLE 2. Treatment coding agreement by the kappa statistic, by database

	Hosp†	Yes	Yes	No	No	
Treatment	**MCR†**	**Yes**	**No**	**Yes**	**No**	**Kappa**
BCS, Definition 1[3]		1,246	23	474	2,213	0.74
BCS, Definition, 2[4]		1,600	57	120	2,179	0.91
Mastectomy		1,969	45	39	1,903	0.96
AND		2,538	36	368	1,014	0.76

	Hosp†	Yes	Yes	No	No	
Treatment	**Med†**	**Yes**	**No**	**Yes**	**No**	**Kappa**
BCS, Definition, 1[3]		NA	NA	NA	NA	NA
BCS, Definition, 2[4]		NA	NA	NA	NA	NA
Mastectomy		1,993	21	343	1,599	0.82
AND		2,507	67	360	1,022	0.75

	Med†	Yes	Yes	No	No	
Treatment	**MCR†**	**Yes**	**No**	**Yes**	**No**	**Kappa**
Breast Conservation[1]		NA	NA	NA	NA	NA
Breast Conservation[2]		NA	NA	NA	NA	NA
Mastectomy		1,991	345	17	1,603	0.82
AND		2,796	71	110	979	0.88

† Hosp: Hospital records, MCR: Manitoba Cancer Registry, Med: Medical claims
NA: Not available
[3] ICES definition – Codes 85.21–85.23
[4] Current Study definition – Codes 85.21–85.23 and 85.12

Procedure rates by data set are shown in Table 3. Using different sources of treatment information produced somewhat different estimates of treatment prevalence. To assess the healthcare system's treatment of breast cancer patients in our jurisdiction, we also report rates of primary breast cancer surgery, i.e., all women receiving either BCS or mastectomy. Primary breast cancer surgery was not performed in 5.8% of patients in the MCR; the majority of these patients had advanced (Stage IV) disease at diagnosis.

TABLE 3. Surgical procedure rates by data source

	Manitoba Cancer Registry		Medical Claims File		Hospital Discharge File	
	N	%	*N*	%	*N*	%
BCS[3]	1720	43.5	NA	NA	1269	32.1
BCS[4]	1720	43.5	NA	NA	1657	41.9
Mastectomy	2008	50.8	2336	59.0	2014	50.9
Surgery in the Breast (BCS[3]+Mastectomy) (BCS[4]+Mastectomy)	 3728 3728	 94.2 94.2	 NA NA	 NA NA	 3283 3671	 83.0 92.8
AND	2906	73.5	2867	72.5	2574	65.1

NA: Not available
[3] ICES definition – Codes 85.21–85.23
[4] Current study definition – Codes 85.21–85.23 and 85.12

Discussion

This investigation was performed as part of a larger population-based study designed to look at variations in patterns of breast cancer care. Since clinical acceptance of results rests heavily on the ability to identify breast cancer treatment accurately in the population, it was imperative that the strengths and limitations of our data sources be understood. While most studies examining breast cancer treatment patterns utilize only one data source (Malin et al. 2002b), this study employed multiple sources to explore the accuracy of surgical treatment information for breast cancer patients. Because we found that the MCR provides consistently accurate surgical treatment information for all procedures examined, future work exploring variations in patterns of care in Manitoba will focus on this data source.

Treatment information is often captured in administrative databases. However, when used alone, these files may not capture all patients in the region with the cancer of interest (Malin et al. 2002b). This study provides evidence that capture of breast cancer patients using administrative data sets alone can be quite good; more than 95% of breast cancer patients found in the MCR had treatments recorded in the hospital discharge and the medical claims files; more than 99% of these patients were found to have a breast cancer diagnosis coded in the administrative records. Our linkage rates are consistent with other studies that have found that 80% to 95% of women with known cancers have records in administrative data sets (Pinfold et al. 2000; Ayanian et al. 1993; Potosky et al. 1993). However, this finding does not ensure that everyone with breast cancer recorded in administrative data is found to have breast cancer in the

MCR; our registrars examine many reports that have cancer diagnoses assigned on an interim basis that are ruled out on closer investigation.

A challenge to the accurate reporting of cancer surgery involved identifying data on primary breast procedures (BCS and mastectomy). Other researchers may also find that data sources – even if they cover the entire population – are not equal in their ability to report treatment comprehensively, owing to coding limitations. Our findings indicate that healthcare management agencies must take care to include appropriate activity codes in a timely fashion when new technologies are introduced, and that analysts must take care in understanding the underlying accounting nature of the data system when they use billing data for research.

We were able to confirm the accuracy of several data sources and resolve discrepancies through targeted chart reviews with relatively little effort, considering the thousands of patients and procedures included in the analysis. We have also shown that the population-based cancer registry proved to be superior to other administrative data sets for capturing surgical treatment information on cancer cases. More broadly, we have demonstrated that using different data sets can result in rates for breast surgery that are sufficiently disparate as to warrant some concern and that certain data sources will accurately reflect one procedure while being inaccurate on other procedures. This study illustrates the importance of critically examining and evaluating data sources in health services research in order to select those that will be most appropriate and accurate for the treatments being studied. Care should be taken in the interpretation of results of health services research if the accuracy of the information has not been ascertained.

… the population-based cancer registry proved to be superior to other administrative data sets for capturing surgical treatment information on cancer cases.

Correspondence may be directed to: Donna Turner, Epidemiology & Cancer Registry, CancerCare Manitoba, 675 McDermot Ave., Winnipeg, MB Canada R3E 0V9. Tel.: 204-787-2573. E-mail: Donna.turner@cancercare.mb.ca.

REFERENCES

Ayanian, J.Z., B.A. Kohler, T. Abe and A.M. Epstein. 1993. "The Relation between Health Insurance Coverage and Clinical Outcomes among Women with Breast Cancer." *New England Journal of Medicine* 329(5): 326–31.

Iscoe, N., T. To, E. Gort and M. Tran. 1997. *Cancer Surgery in Ontario.* Toronto: Institute for Clinical Evaluative Sciences.

Landis, J.R. and G.G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33(1): 159–74.

Malin, J.L., K.L. Kahn, J. Adams, L. Kwan, M. Laouri and P.A. Ganz. 2002a. "Validity of Cancer Registry Data for Measuring the Quality of Breast Cancer Care." *Journal of the National Cancer Institute* 94(11): 835–44.

Malin, J.L., M.A. Schuster, K.A. Kahn and R.H. Brook. 2002b. "Quality of Breast Cancer Care: What Do We Know?" *Journal of Clinical Oncology* 20(21): 4381–93.

Middleton, R.J., A.T. Gavin, J.S. Reid and D. O'Reilly. 2000. "Accuracy of Hospital Discharge Data for Cancer Registration and Epidemiological Research in Northern Ireland." *Cancer Causes and Control* 11(10): 899–905.

Pinfold, S.P., V. Goel and C. Sawka. 2000. "Quality of Hospital Discharge and Physician Data for Type of Breast Cancer Surgery." *Medical Care* 38(1): 99–107.

Potosky, A.L., G.F. Riley, J.D. Lubitz, R.M. Mentnech and L.G. Kessler. 1993. "Potential for Cancer Related Health Services Research Using a Linked Medicare-Tumor Registry Database." *Medical Care* 31(8): 732–48.

Call to Authors

Data Matters presents brief, focused papers that report analyses of health administrative or survey data that shed light on significant health services and policy issues. Submissions to Data Matters should be a maximum of 1,500 words, exclusive of abstract (max. 100 words), tables, figures and references, and should include no more than three tables or figures.

Appel aux auteurs

« Questions de données » présente de brefs articles ciblés portant sur des analyses de données administratives ou d'enquêtes sur la santé ou de données d'enquête et qui font la lumière sur d'importantes questions liées aux services et aux politiques de santé. Les articles soumis à « Questions de données » doivent être d'au plus 1 500 mots, excluant le résumé (100 mots au plus), les tableaux, diagrammes et références et ne doivent pas comprendre plus de trois tableaux ou diagrammes.

For more information contact Rebecca Hart, Managing Editor, at rhart@longwoods.com.

Pour de plus amples renseignements, veuillez communiquer avec Rebecca Hart, rédactrice, á rhart@longwoods.com.