Journal of Gastrointestinal Surgery (2019) 23:367–376
https://doi.org/10.1007/s11605-018-4043-z

ORIGINAL ARTICLE

Do Diagnostic and Procedure Codes Within Population-Based, Administrative Datasets Accurately Identify Patients with Rectal Cancer?

Reilly P. Musselman[1] · Tara Gomes[2] · Deanna M. Rothwell[3,4] · Rebecca C. Auer[1] · Husein Moloo[1] · Robin P. Boushey[1] · Carl van Walraven[2,3]

Received: 27 August 2018 / Accepted: 29 October 2018 / Published online: 3 December 2018

Abstract

Background Procedural and diagnostic codes may inaccurately identify specific patient populations within administrative datasets.

Purpose Measure the accuracy of previously used coding algorithms using administrative data to identify patients with rectal cancer resections (RCR).

Methods Using a previously published coding algorithm, we re-created a RCR cohort within administrative databases, limiting the search to a single institution. The accuracy of this cohort was determined against a gold standard reference population. A systematic review of the literature was then performed to identify studies that use similar coding methods to identify RCR cohorts and whether or not they comment on accuracy.

Results Over the course of the study period, there were 664,075 hospitalizations at our institution. Previously used coding algorithms identified 1131 RCRs (administrative data incidence 1.70 per 1000 hospitalizations). The gold standard reference population was 821 RCR over the same period (1.24 per 1000 hospitalizations). Administrative data methods yielded a RCR cohort of moderate accuracy (sensitivity 89.5%, specificity 99.9%) and poor positive predictive value (64.9%). Literature search identified 18 studies that utilized similar coding methods to derive a RCR cohort. Only 1/18 (5.6%) reported on the accuracy of their study cohort.

Conclusions The use of diagnostic and procedure codes to identify RCR within administrative datasets may be subject to misclassification bias because of low PPV. This underscores the importance of reporting on the accuracy of RCR cohorts derived within population-based datasets.

Keywords Rectal cancer · Administrative data

Introduction

With increased digitization of health care and enhanced computing power, administrative data have become an increasingly available resource for clinical researchers. The use of administrative databases for health care-related research publications in surgery has risen steadily over the past decade, with results that often report on important clinical outcomes.[1–3] The majority of administrative database research uses diagnostic or procedural codes to identify diseases, exposures, or outcomes.[4]

Several problems exist regarding the use of health care-related codes to identify patient cohorts within administrative datasets. The accuracy of administrative data codes used to identify different diseases varies widely.[5,6] There can also be extensive variability between studies in the numbers and types of codes used to identify the same population.[7] Additionally, very few studies actually validate the codes (or coding algorithms) they use to create datasets for analysis.[4] This indicates that the accuracy of the surrogate marker (the code) for the entity it is supposed to represent is unknown. As such, the

✉ Reilly P. Musselman
rmusselman@toh.ca

[1] Division of General Surgery, University of Ottawa, Ottawa, ON, Canada

[2] Institite for Clinical and Evaluative Sciences, Toronto, ON, Canada

[3] Ottawa Hospital Research Institute, Ottawa, ON, Canada

[4] The Ottawa Hospital, Ottawa, ON, Canada

generalization of results using the code to those involving the entity that is trying to be studied is unreliable.[8]

Traditional procedure codes for rectal resections include those for anterior resections (AR), low anterior resections (LAR), and abdominal perineal resections (APR). In order to identify patients who have undergone rectal cancer resection (RCR), researchers typically will combine these procedure codes with a diagnosis code consistent with either rectal or colorectal cancer.[9] In addition to the inherent inaccuracies of procedure and diagnostic codes, it is especially difficult to differentiate cancer of the rectum from cancer of the sigmoid colon by simply using these codes alone. That is, even if the codes are accurate, they do not themselves pinpoint the location of the tumor since they are commonly used interchangeably for surgical resection of both sigmoid and rectal cancers. Therefore, accurately identifying patients with RCR and distinguishing them from colon cancer patients within large, population-based, administrative datasets can be problematic. This is important because rectal cancer differs from colon cancer with respect to workup, treatment algorithms, and outcomes.[10–13]

Before we can confidently use population-based administrative datasets to study clinical outcomes for patients with RCR, we need to be sure that the patient cohort derived from these databases accurately reflects the patient population we wish to study. The purpose of the present study was to measure the accuracy of a cohort of RCRs that was derived using diagnostic and procedure codes within population-based administrative databases and published in a peer-reviewed journal.

Materials and Methods

Sources of Data

The Ottawa Hospital Data Warehouse (OHDW) is a relational database containing administrative and clinical data from all operational information systems at The Ottawa Hospital (TOH), a tertiary care institution. Operational information systems include clinical data such as laboratory results, some clinical notes, pathology reports, hospitalization discharge abstract records, and demographic data.

The Institute for Clinical and Evaluative Sciences (ICES) provides linked access to several provincial, population-based administrative datasets. The present study used ICES to obtain patient records from the Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD; captures administrative, clinical, and demographic information for all hospital discharges, including deaths, in the province of Ontario), the Registered Persons Database (RPD; contains demographic data for all people with an Ontario Health Card Number), and the Ontario Cancer Registry (OCR; includes information on all newly diagnosed cases of invasive neoplasia in Ontario).

Rectal Cancer Cohort

Using previously published methods,[14] we identified all patients with RCR at a tertiary care facility between April 1st, 2002 and December 31st, 2010. These methods utilized text-search of pathology reports which were shown to be 100% sensitive and 98.6% specific.[14] In brief, all pathology reports were manually reviewed by RM to identify potential clauses and phrases that were repeated frequently and could be used in a text search algorithm to identify other, similar pathology reports of RCRs. Using SAS statistical software, version 9.2 (SAS Institute Inc., Cary NC), we applied these phrases to a coding macro that analyzes text for clauses of interest with or without preceding or following specified qualifiers and applied this macro to all pathology reports at TOH over the course of the study period. False positive records identified through text-search methods were later excluded through manual chart review to identify all relevant cases. The resultant patient cohort was used as the gold standard population of all RCRs performed at The Ottawa Hospital over the course of the study period. Therefore, all patients who were not in this *rectal cancer cohort* were deemed not to have had surgically resected rectal cancer.

Study Cohort

We previously published a study on RCRs within ICES databases that meets criteria for the literature search described above.[9] In this study, all inpatient encounters performed in the province of Ontario between April 1st 2002 and December 31st, 2010 were identified within CIHI-DAD (Fig. Fig. 1a). Using a unique patient identifier, this cohort was linked to the Ontario Cancer Registry to identify the patients with a pathological diagnosis of colorectal cancer in OCR within 1 year before or 28 days after the hospital encounter date (Fig. Fig. 1b). Canadian Classification of Health Intervention (CCI) procedure codes were then used to determine which of these patients had an anterior resection, low anterior resection, or abdominal-perineal resection (Fig. Fig. 1c) (Appendix Table 3). The resultant cohort was thought to represent all patients in the province of Ontario with RCR between April 1st, 2002 and December 31st, 2010 (Fig. Fig. 1d).

The above coding algorithm was repeated in an identical manner, with results limited to a single, tertiary-care institution and tested for its overall accuracy. Accuracy statistics for the coding algorithm were calculated using the *rectal cancer cohort* as the reference standard. All statistics will be compiled and analyzed using SAS statistical software, version 9.2 (SAS Inc., Cary, North Carolina).

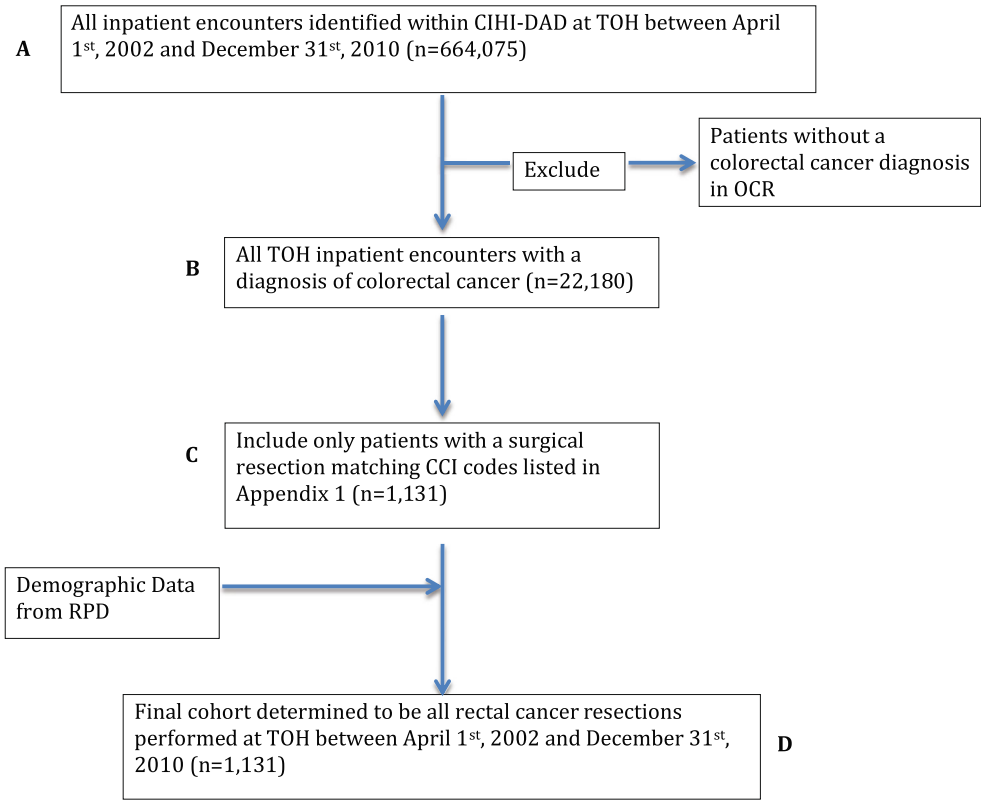

Fig. 1 Flow diagram illustrating methodology used in previously published studies to identify rectal cancer resections through administrative databases (re-created with permission from the authors[15]). TOH = The Ottawa Hospital; CIHI-DAD = Canadian Institute for Health Information, Discharge Abstract Database; CCI = Canadian Classification of Health Interventions; OCR = Ontario Cancer Registry; RPD = Registered Persons Database

Literature Search

We performed a literature review by searching MEDLINE and EMBASE databases for all studies from 2007 onwards that used administrative data to identify patients with RCR. Specific search terms as well as the search strategy are outlined in Appendix Table 4. Studies were included if they used a combination of diagnosis and procedure codes to identify their patient cohort within a population-based, administrative dataset. Studies were excluded if they used a cancer registry as the sole source of data collection, used a prospectively collected dataset, or did not differentiate between colon and rectal resections within the results. For all included studies, we recorded the clinical outcomes reported by the paper and whether or not the authors commented on the validity of the codes used to identify their cohort, or the accuracy of the cohort itself within the text of the paper.

Results

Rectal Cancer Cohort

A total of 664,075 hospitalizations occurred at our institution between April 1st, 2002 and December 31st, 2010 (inclusive). A total of 821 patients had a RCR during the study period and were successfully linked to administrative records at the Institute for Clinical and Evaluative Sciences (ICES) servers. Sixty-five records failed to link and were excluded from analysis because of missing patient unique identifiers, missing or invalid admission dates, or missing records in the Discharge Abstract Database (DAD) (Fig. 2). This represents a RCR incidence of 1.24 per 1000 hospitalizations.

Study Cohort Validation

A total of 664,075 records were identified through the DAD in ICES for patients 18 years or older who were admitted to The Ottawa Hospital (TOH) between April 1st, 2002 and December 31st, 2010 (the same time period of the *rectal cancer cohort*). A total of 22,180 of these patients (3.34%) had a diagnostic code for colorectal cancer in the Ontario Cancer Registry between 1 year prior to, or up to 28 days after, their admission date. A total of 1131 of these patients (5.1%) had at least one CCI procedure code in a DAD record that matched those listed in Appendix Table 3. This resulted in a study cohort of 1131 RCRs identified through ICES datasets using previously described methods (incidence of 1.70 per 1000 hospitalizations).

Using the *rectal cancer cohort* as the reference standard, sensitivity and specificity of the *study cohort* were 89.5% and 99.9%, respectively (Table 1). The prevalence of true rectal

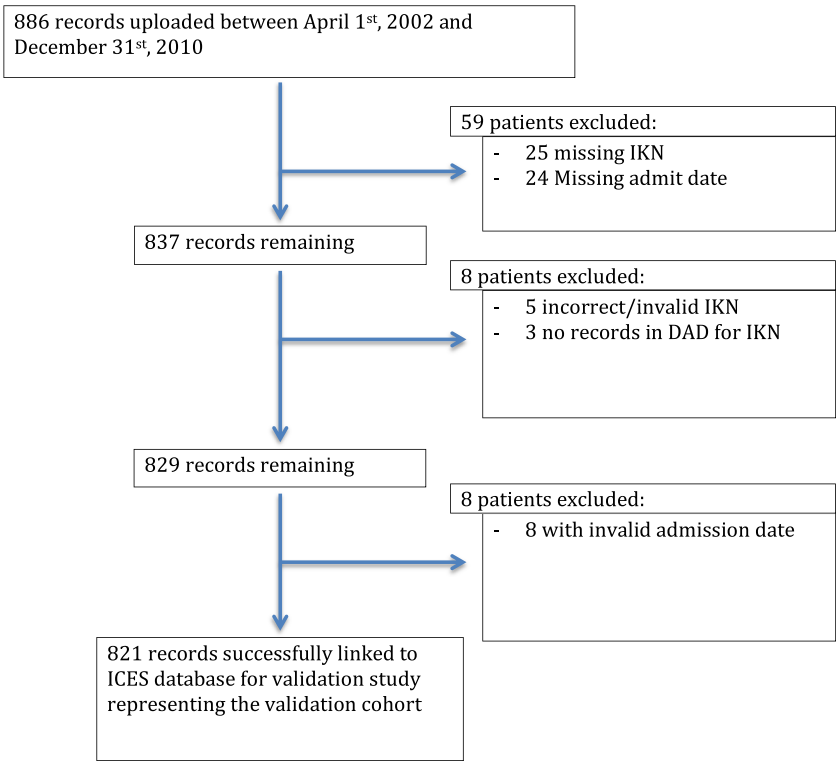

Fig. 2 Flow diagram representing the linkage of the gold standard *validation cohort* onto ICES servers for comparison with the *study cohort*. IKN = ICES Key Number. ICES = Institute for Clinical and Evaluative Sciences. DAD = Discharge Abstract Database

cancer resections in this population was very low at 0.124% of admissions, resulting in a positive likelihood ratio (LR+) of 1499.44. There were 396 false positives captured through the study cohort resulting in a positive predictive value (PPV) of the algorithm was 64.9% (Table 1).

Table 1 2 × 2 table demonstrating sensitivity, specificity, PPV, and NPV of previously published coding algorithm for identifying rectal cancer resections in ICES when compared to a validation cohort (N = 664,075)

		Validation cohort[a]		Totals
		+	–	
Study cohort[b]	+	***A*** 735	***B*** 396	1131
	–	***C*** 86	***D*** 662,858	662,944
	Totals	821	663,254	664,075

Sensitivity = A/A = C = 735/821 = 89.5%

Specificity = D/B + D = 662,858/663,254 = 99.9%

PPV = A/A + B = 735/1131 = 64.9%

NPV = D/C + D = 662,858/662,944 = 99.9%

[a] Validation cohort refers to the validated cohort of rectal cancer patients derived and validated through manual chart review in a prior study

[b] Study cohort refers to the rectal cancer resections identified through previously published coding algorithms mining ICES databases

PPV positive predictive value, *NPV* negative predictive value, *ICES* Institute for Clinical and Evaluative Sciences

Literature Search

We identified 1570 journal articles using the search strategy outlined in Appendix Table 4. From these, 1321 were excluded based on review of abstracts by RM and a further 231 were excluded after review of full text manuscripts for various reasons. The final result was 18 studies that used a combination of diagnostic and procedure codes to identify patients who had surgical resection for rectal cancer within population-based administrative datasets (Fig. 3).[9,16–32]

These studies used two primary methods for identifying their patient cohort. Five of the 18 studies (27.8%) used diagnostic codes within a dedicated, population-based cancer registry to confirm a diagnosis of either rectal or colorectal cancer and linked this cohort to a population-based administrative dataset to capture information on hospitalizations, including surgical procedures (Table 2). The remaining 13 studies (72.2%) used diagnostic and procedure codes within a single, population-based, administrative dataset to identify RCRs (Table 2). In both cases, an index rectal resection was identified using procedure codes. This index procedure was subsequently defined as an RCR if an associated diagnostic code for rectal or colorectal cancer was present within a specified time-frame of the index procedure.

One out of 18 studies (5.6%) validated their patient cohort through manual chart review. Of the remaining 17 studies, none commented on the overall accuracy of the patient cohort, while 2/17 (11.7%) commented on the overall accuracy of the codes used within their coding algorithm. Outcomes reported in these studies included utilization rates for laparoscopic surgery, rates of

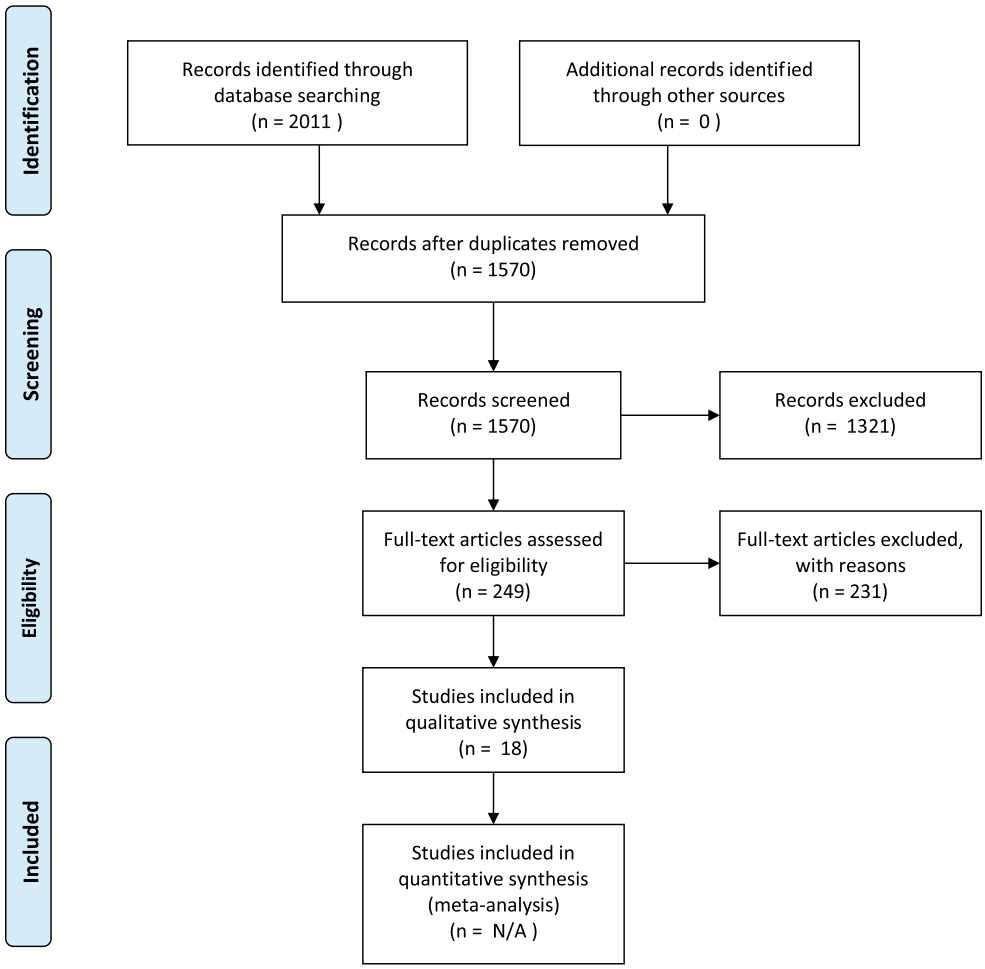

Fig. 3 Flow diagram illustrating literature review of MEDLINE and EMBASE databases resulting in 18 included studies

abdominoperineal resections (APR), in-hospital morbidity and mortality, cancer-specific survival, and overall survival (Table 2).

Discussion

We previously reported on RCR trends on a population level, defining the RCR cohort using diagnostic in procedure codes.[9] In the present study, we tested the accuracy of our methods against a validated patient cohort of RCRs and found adequate sensitivity and specificity, but a low PPV. Our results are therefore susceptible to bias due to patient misclassification since roughly 35% of patients in the denominator did not have the entity being studied.

In a systematic review of the literature, we identified 18 studies that used a similar combination of diagnostic and procedure codes to identify patients with RCR within large, population-based administrative datasets. Many of these studies reported on important clinic outcomes and were published in high-impact surgical journals. However, very few commented on the validity of the codes used to define the RCR cohort, or the overall accuracy of the cohort itself. As a result, they may be susceptible to similar mis-classification bias.

PPV varies depending on disease prevalence. In our study, only 821 of 664,075 hospitalizations were for RCR, yielding a disease prevalence of 0.12%. With such a low pre-test probability, an inadequate PPV can occur despite adequate sensitivity and specificity. The nature of population-based administrative data research is such that a relatively small number of specific disease processes or outcomes of interest are extracted from hundreds of thousands, and sometimes millions, of health-care encounters. By default, almost any entity being studied will have a low prevalence within these datasets. Therefore, even if the codes used to extract these data have adequate accuracy statistics, it is still possible for the overall coding algorithm to have a low PPV. This underscores the importance of validating the methods used to define patient cohorts within these large datasets.

The accuracy of coding algorithms to identify patient cohorts using administrative data varies widely among studies and disease sites. Widdifield et al. tested over 100 coding algorithms used to identify patients with rheumatoid arthritis and found that the PPV ranged between 51 and 83%.[33] Another study that used procedure codes to identify patients from the CIHI discharge abstract database (DAD) in Canada demonstrated PPVs of 95%, 96%, and 98% for percutaneous coronary intervention (PCI), cardiac catheterization, and coronary artery bypass graft

Table 2 Results of review of the literature to identify studies that used diagnostic and procedure codes to create a cohort of patients' rectal cancer resections within a population-based, administrative database

Study	Journal	Database type	Reported outcomes	Diagnosis codes	Procedure codes	Validation reported	
						Codes	Cohort
Ackerman et al.[16]	*Journal of Medical Economics*	DAD	Utilization[a], short-term morbidity	ICD-9	ICD-9-CM	No	No
Saia et al.[17]	*Updates in Surgery*	DAD	Utilization	ICD-9	ICD-9	No	No
Yeo et al.[18]	*Annals of Surgery*	DAD	Short-term morbidity, surgeon volumes	ICD-9-CM	ICD-9-CM	No	No
Keller et al.[19]	*Surgical Endoscopy*	DAD	Utilization	ICD-9	ICD-9	No	No
Aquina et al.[20]	*Surgery*	DAD	APR proportion, surgeon volume	ICD-9	ICD-9	No	No
Wiegard et al.[21]	*Annals of Surgical Oncolology*	DAD	Short-term morbidity	ICD-9	ICD-9	No	No
Dobbins et al.[22]	*Diseases of the Colon & Rectum*	DAD, CR	Utilization, readmission rates	ICD-O-3	ICD-10	No	No
Simunovic et al.[23]	*Annals of Surgical Oncology*	DAD, CR	Utilization, short-term mortality, cancer recurrence	NS	NS	Yes	No
Musselman et al.[9]	*Colorectal Disease*	DAD, CR	Utilization	ICD-9	ICD-10	No	No
Devon et al.[24]	*Surgery*	DAD, CR	Utilization, readmission rates	ICD-9	ICD-9	Yes	No
Ricciardi et al.[25]	*Diseases of the Colon & Rectum*	DAD	APR proportion	ICD-9	ICD-9	No	No
Ricciardi et al.[26]	*Diseases of the Colon & Rectum*	DAD	Colostomy rates, surgeon specialty	ICD-9	ICD-9	No	No
Ricciardi et al.[27]	*Diseases of the Colon & Rectum*	DAD	APR proportion	ICD-9	ICD-9	No	No
Thompson et al.[28]	*Medical Journal of Australia*	DAD	Utilization	ICD-10	ICD-10	No	No
Tilney et al.[29]	*Annals of Surgery*	DAD	APR proportion	ICD-9	OPCS4	No	No
Singh et al.[30]	*Surgical Oncology*	DAD	Short-term morbidity	ICD-9-CM	ICD-9-CM	No	Yes
Ricciardi et al.[31]	*Diseases of the Colon & Rectum*	DAD	Predictors for sphincter sparing surgery	ICD-9	ICD-9	No	No
Morris et al.[32]	*GUT*	DAD, CR	APR proportion	ICD-9	OPCS4	No	No

[a] Utilization refers to resource utilization for different operative approaches such as laparoscopic, open, and robotic

DAD Discharge Abstract Database, *CR* Cancer Registry, *APR* abdominal perineal resection, *ICD* International Classification of Disease, *OPCS* Office of Population Censuses and Surveys

(CABG) surgery, respectively.[34] It is unclear what is the acceptable level for PPV when deriving patient cohorts. For example, Tu et al. used a coding algorithm with a PPV of 71% to identify patients with atrial fibrillation[35]. While the authors recognize that this is on the low end for PPV, they still use the cohort to measure percentage of people receiving anticoagulation.

Little has been reported on the accuracy of administrative datasets for identifying patients who have undergone rectal cancer surgery. Goldsbury et al. reported the accuracy for identifying rectal cancer surgery within a population-based cancer registry in New South Wales, Australia. They demonstrated a sensitivity and specificity of 85% and 96%, respectively, which are similar to our results.[36] However, this study collected those data from a prospectively collected cancer registry whose sole purpose is to record demographic, treatment, and outcomes data for cancer patients within a population. Such a registry is distinctly different from administrative data, which is defined as data collected and abstracted for administrative, and not clinical, purposes.[37] Li and colleagues reported an algorithm for identifying rectal cancer surgeries in administrative datasets with a PPV of 93%, much higher than our results.[38] However, this was also from a provincial cancer registry, which carries a much higher pre-test probability for rectal cancer than other, population-based datasets.

An alternative to using codes within administrative databases is text-search methodology that utilizes natural language processing (NLP) or artificial intelligence (AI) to study clinical records such as pathology reports, radiology reports, or clinic notes. These techniques have demonstrated potential for improved discrimination within clinical reports, which may become more relevant with the widespread adoption of electronic medical records (EMR).[39] Using advanced technology, however, may come with increased costs and longer timeframes. Regardless of the methods used to identify patient cohorts within large, population-based datasets, determining and reporting the accuracy of your study cohort are important prior drawing any clinically relevant conclusions.

There are several limitations to our study. First, we used a single institution to validate the previously described coding algorithm. The algorithm that we tested attempted to identify all rectal cancer resections in the province of Ontario using administrative datasets available through ICES.[9] While the present study validated the accuracy of this algorithm, we did so using a gold standard cohort from only one institution within the province. This can limit the generalizability of the results. Other institutions that perform rectal cancer surgery in Ontario may have more or less reliable data entry methods into administrative datasets such as CIHI. A more complete validation study would therefore utilize at least a sample of rectal cancer surgery cases from multiple institutions around the province, thereby making the results more generalizable.

Second, we only validated the RCR from a single study to obtain accuracy statistics of the methods used to derive the patient cohort. Although we were able to identify 18 other studies in the literature that use similar methods, we did not have access to the exact coding algorithm used in these papers. The overall accuracy of these patient cohorts may be more or less accurate than the one validated in the present study. Despite this limitation, we maintain that our results demonstrate the need to report on the accuracy of RCR cohorts derived using administrative data, which is rarely done.

Conversely, a strength of our study is that it validates the complete methodology used to create the patient cohort, rather than just the codes themselves. The CCI codes used in ICES databases, specifically CIHI-DAD, to identify colorectal surgical resections have been validated against chart extraction, and were found to be very reliable with a sensitivity of 94% and a PPV of 90%.[15] However, simply using validated procedure or diagnostic codes does not necessarily ensure that an entire algorithm used to extract data from large, administrative datasets is also valid, or that the final cohort will be accurate. The present study examines the actual accuracy of the cohort that was created, providing a probability that a patient identified actually has the disease in question. This has a much more practical application for researchers who use administrative data, providing a reliable method for identifying a specific patient cohort.

In summary, when trying to identify patients with RCR within population-based, administrative datasets, traditional diagnostic and procedure codes yield adequate sensitivity and specificity, but potentially low positive predictive value. This is primarily due to the low pre-test probability of having the disease in question that is inherent to population-based data. This study confirms the importance of validating coding algorithms used to derive RCR cohorts and outcomes within administrative datasets. Future research should focus on minimizing the misclassification bias in administrative database research by validating and improving the accuracy of these large patient cohorts.

Authors' Contributions All authors contributed to the concept and design of the study, as well as drafting and revising of the manuscript. All authors gave final approval for the manuscript. Authors Musselman, Gomes, Rothwell, Auer, and vanWalraven were directly involved in data collection and analysis.

Funding This study was supported by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. Parts of this material are based on data and information compiled and provided by Canadian Institute for Health Information (CIHI). However, the analyses, conclusions, opinions, and statements expressed herein are those of the author, and not necessarily those of CIHI.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Appendix

Table 3 CCI codes used to define cohort

Code	Description
1NQ87DF	Excision partial, rectum using laparoscopic approach with enterocolostomy
1NQ89GV	Excision total, rectum using combined laparoscopic with perineal approach with coloanal or ileoanal anastomosis
1NQ87PB	Excision partial, rectum using open approach with colorectal anastamosis
1NQ87RD	Excision partial, rectum using open approach with colorectal or coloanal anastamosis
1NQ89KZ	Excision total, rectum using open abdominoperineal approach
1NQ89KZXX	Excision total, rectum using abdominoperineal approach pouch formation with ileoanal anastamosis
1NQ89SF	Excision total, recum using open approach with bypass technique
1NQ89SFXXG	Excision total, rectum abdominal approach and pouch formation
1NQ90LAXXG	Excision total with reconstruction, rectum using open approach with ileoanal anastamosis
1NQ87DE	Excision partial, rectum using laparoscopic approach with colorectal anastamosis
1NQ89LH	Excision total, rectum using open abdominoperineal approach
1NQ89LHXXG	Excision total, rectum using abdominoperineal approach and continent ileostomy formation
1NQ89RS	Excision total, rectum using open anterior approach with terminal colostomy
1NQ89RSXXG	Excision total, rectum using abdominal approach and continent ileostomy formation
1NQ89AB	Excision total, rectum using combined approach, laparoscopic, stoma formation with distal closure
1NQ87DX	Excision partial, rectum using laparoscopic approach, stoma formation and distal closure
1NQ87TF	Excision partial, rectum using open approach, stoma formation and distal closure

CCI Canadian Classification of health Interventions

Table 4 Search strategy used for MEDLINE and EMBASE literature search

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) < 1946 to Present>	
Search strategy:	
1	"International Classification of Diseases"/ (6977)
2	"international classification of disease".tw. (1159)
3	(icd9* or icd10* or icd).tw. (29130)
4	REGISTRIES/ (72746)
5	(registry or registries).tw. (97633)
6	exp Medical Records/ or (EHR or EMR or EPR).tw. (156879)
7	(utili?ation data* or claims data* or managed care data* or billing data* or hospitali?ation data* or linked data*:).tw. (14572)
8	((health or patient or medical) adj2 record*).tw. (115001)
9	(physician* adj2 claim*).tw. (683)
10	((clinic* or patient) adj2 (code or coding)).tw. (928)
11	(admin* adj2 bill*).tw. (137)
12	Databases, Factual/ (63561)
13	(admin* adj3 data*).tw. (18324)
14	or/1–13 (475430)
15	exp Rectal Neoplasms/su [Surgery] (17965)
16	((rectal or colorectal or colo-rectal) adj2 (cancer or neoplasm* or lesion* or adenocarcinoma* or tumor* or tumour*) adj4 (surg* or resection*)).tw. (11672)
17	15 or 16 (25310)
18	14 and 17 (1147)
19	limit 18 to yr = "2007 -Current" (866)
Database: Embase Classic + Embase < 1947 to 2018 March 02>	
Search strategy:	
1	*"International Classification of Diseases"/ (1794)
2	"international classification of disease".tw. (1837)
3	(icd9* or icd10* or icd).tw. (62918)
4	*register/ (24500)
5	(registry or registries).tw. (160923)
6	*medical record/ (33913)
7	(utili?ation data* or claims data* or managed care data* or billing data* or hospitali?ation data* or linked data*:).tw. (25037)
8	((health or patient or medical) adj2 record*).tw. (183779)
9	(physician* adj2 claim*).tw. (980)
10	((clinic* or patient) adj2 (code or coding)).tw. (1587)
11	(admin* adj2 bill*).tw. (244)
12	*factual database/ (5952)
13	(admin* adj3 data*).tw. (27334)
14	or/1–13 (465823)
15	exp *rectum cancer/ and cancer surgery/ (13203)
16	((rectal or colorectal or colo-rectal) adj2 (cancer or neoplasm* or lesion* or adenocarcinoma* or tumor* or tumour*) adj4 (surg* or resection*)).tw. (17782)
17	15 or 16 (26093)
18	14 and 17 (1318)
19	limit 18 to yr = "2007 -Current" (1145)

References

1. Rodriguez B, Balla A, Fernandez-Ananin S, Balague C, Targarona EM. The era of large databases: outcomes after gastroesophageal surgery according to NSQIP, NIS and NCDB databases. Systematic literature review. *Surg Innov.* 2018; 25(4):400–412.
2. Rodriguez B, Balla A, Corradetti S et al. What have we learned in minimally invasive surgery from NSQIP and NIS large databases? A systematic review. *Int J Colorectal Dis.* 2018; 33(6): 663–681
3. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg,* 2015; 102(2): e93-e101.
4. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently uses validated diagnostic or procedural codes. *J Clin Epidemiol*. 2011; 64:1054–1059.
5. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med* 2001;23:205–211.
6. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42:1066–1072.
7. Hohl CM, Karpov AK, Reddekopp L, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. *J Am Med Inform Assoc* 2014; 21:547–557.
8. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in Medicine 1989, v8 n4, 431–440.
9. Musselman RP, Gomes T, Chan BP, Auer RC, Moloo H, Poulin EC, Mamazza J, Al-Khayal K, Mamdani M, Al-Omran M, Al-Obaid O, Boushey RP. Changing Trends in Rectal Cancer Surgery in Ontario: 2002-2009. *Colorectal Disease*. 2012; 14(12):1467–1472.
10. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on cancer sixth edition staging. *Journal of the National Cancer Institute.* 2004; 96(19): 1420–1425.
11. AJCC (American Joint Committee on Cancer) Cancer Staging Manual, 7th edition, Edge, SB, Byrd, DR, Compton, CC, et al (Eds), Springer, New York 2010. p.143
12. Wong R, Berry S, Spithoff K et al. Preoperative or Postoperative Therapy for the Management of Patients with Stage II or III rectal cancer: Guideline Recommendations. July, 2008. Available at https://www.cancercare.on.ca/cms/One.aspx?portalId=1377&pageId=10207. Accessed 08/07, 2013.
13. Tjandra JJ, Kilkenny JW, Buie WD et al. Practice parameters for the management of rectal cancer (revised). *Dis Colon rectum*. 2005; 48: 411–423.
14. Musselman RP, Rothwell D, Auer RC, Moloo H, Boushey R, van Walraven C. Identifying rectal cancer patients within administrative datasets using pathology reports. *J Pathol Inform* 2018; 9:18 (ePub ahead of print).
15. Juurlink D, Preyra C, Croxford R et al. Canadian Institute for Health Information Discharge Abstract Database: a validation study. Toronto: *Institute for Clinical Evaluative Sciences*; 2006.
16. Ackerman SJ, Shoshana D, Baik R et al. Comparison of complication and conversion rates between robotic-assisted and laparoscopic rectal resection for rectal cancer: which patients and providers could benefit most from robotic-assisted surgery?. *J Med Econ.* 2018; 21(3): 254–261
17. Saia M, Buja A, Mantoan D et al. Isolated rectal cancer surgery: a 2007-2014 population study based on a large administrative database. *Updates Surg.* 2017; 69: 367–373
18. Yeo HL, Abelson JS, Mao J et al. Surgeon annual and cumulative volumes predict early postoperative outcomes after rectal cancer resection. *Ann Surg.* 2017;265: 151–157
19. Keller D, Qiu J, Senagore AJ. Predicting opportunities to increase utiliziation of laparoscopy for rectal cancer. *Surg Endosc.* 2018; 32: 1556–1563.
20. Aquina CT, Probst CP, Becerra BA et al. High volume improves outcomes: the argument for centralization of rectal cancer surgery. *Surgery* 2016;159: 736–748.
21. Wiegard NE, Hart KD, Herzig DO. Psychiatric illness is a disparity in the surgical management of rectal cancer. *Ann Surg Oncol* 2015;22: S573-S579.
22. Dobbins TA, Young JM, Solomon MJ. Uptake and outcomes of laparoscopically assisted resection for colon and rectal cancer in Australia: a population-based study. *Dis Colon Rectum* 2014;57: 415–422
23. Simunovic M, Baxter NN, Sutradhar R, Liu N, Cadeddu M, Urbach D. Uptake and patient outcomes of laparoscopic colon and rectal cancer surgery in a publicly funded system and following financial incentives. *Ann Surg Oncol.* 2013; 20: 3740–3746.
24. Devon KM, Urbach DR, McLeod RS. Postoperative disposition and health services use in elderly patients undergoing colorectal cancer surgery: a population-based study. *Surgery* 2011; 149: 705–712.
25. Ricciardi R, Roberts PL, Read TE, Baxter NN, Marcello PW, Schoetz DJ. Who performs proctectomy for rectal cancer in the United States. *Dis Colon Rectum* 2011; 54;1210–1215.
26. Ricciardi R, Roberts PL, Read TE, Baxter NN, Marcello PW, Schoetz DJ. Presence of specialty surgeons reduces the likelihood of colostomy after proctectomy for rectal cancer. *Dis Colon Rectum.* 2011; 54;207–213.
27. Ricciardi R, Roberts PL, Read TE, Marcello PW, Schoetz DJ, Baxter NN. Variability in reconstructive procedures following rectal cancer surgery in the United States. *Dis Colon Rectum* 2010; 53;874–880.
28. Thompson BS, Coory MD, Lumley JW. National trends in the uptake of laparoscopic resection for colorectal cancer, 2000-2008. *MJA* 2011; 194: 443–447
29. Tilney HS, Heriot AG, Purkayastha et al. A national perspective on the decline of abdominoperineal resection for rectal cancer. *Ann Surg.* 2008;247(1): 77–84.
30. Singh RK, Dharmasena D, Virgo K, et al. Proctectomy for rectal cancer in patients with prior spinal cord injury. *Surg Oncol.* 2008;17:313–316.
31. Ricciardi R, Vernig BA, Madoff RD, Rothenberger DA, Baxter NN. The status of radical proctectomy and sphincter sparing surgery in the United States. *Dis Colon Rectum.* 2007; 50: 1119–1127.
32. Morris E, Quirke P, Thomas JD et al. Unacceptable variation in abdominoperineal excision rates for rectal cancer: time to intervene?. *Gut* 2008; 57:1690–1697.
33. Widdifield J, Bombardier C, Bernatsky S et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord*. 2014; 15:216–225.
34. Lee DS, Stitt A, Wang X et al. Administrative hospitalization database validation of cardiac procedure codes. *Med Care*. 2013; 51(4):e22–26.
35. Tu K, Nieuwlaat R, Cheng S et al. Identifying patients with atrial fibrillation in administrative data. *Canadian Journal of Cardiology.* 2016; 32:1561–1565.
36. Goldsbury DE, Armstrong K, Simonella L, Armstrong BK, O'Connell DL. Using administrative health data to describe colorectal and lung cancer in New South Wales, Australia: a validation study. *BMC Health Services Research*. 2012; 12:387–396.
37. The Center for Administrative Data Research. http://cadr.wustl.edu. Updated 2018. Accessed on 05/23, 2018.
38. Li X, King C, deGara C, White J, Winget M. Validation of colorectal cancer surgery data from administrative data sources. *BMC Medical Research Methodology.* 2012; 12:97–104.
39. Savova GK, Masanz JJ, Ogren PV et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J AM Med Inform Assoc 2010; 17(5): 507–513