

Accuracy of administrative claims data for polypectomy

Jonathan M. Wyse MD MSc, Lawrence Joseph PhD, Alan N. Barkun MD MSc, Maida J. Sewitch PhD

ABSTRACT

Background: The frequency of polypectomy is an important indicator of quality assurance for population-based colorectal cancer screening programs. Although administrative databases of physician claims provide population-level data on the performance of polypectomy, the accuracy of the procedure codes has not been examined. We determined the level of agreement between physician claims for polypectomy and documentation of the procedure in endoscopy reports.

Methods: We conducted a retrospective cohort study involving patients aged 50–80 years who underwent colonoscopy at seven study sites in Montréal, Que., between January and March 2007. We obtained data on physician claims for polypectomy from the Régie de l'Assurance Maladie du Québec (RAMQ) database. We evaluated the accuracy of the RAMQ data against information in the endoscopy reports.

Results: We collected data on 689 patients who underwent colonoscopy during the study period. The sensitivity of physician claims for polypectomy in the administrative database was 84.7% (95% confidence interval [CI] 78.6%–89.4%), the specificity was 99.0% (95% CI 97.5%–99.6%), concordance was 95.1% (95% CI 93.1%–96.5%), and the kappa value was 0.87 (95% CI 0.83–0.91).

Interpretation: Despite providing a reasonably accurate estimate of the frequency of polypectomy, physician claims underestimated the number of procedures performed by more than 15%. Such differences could affect conclusions regarding quality assurance if used to evaluate population-based screening programs for colorectal cancer. Even when a high level of accuracy is anticipated, validating physician claims data from administrative databases is recommended.

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to:
Dr. Jonathan Wyse,
jonwyse@gmail.com

CMAJ 2011. DOI:10.1503/cmaj.100897

Population-based screening programs for colorectal cancer rely heavily on the performance of colonoscopy as either the initial examination or as the follow-up to a positive screening by virtual colonography, double-contrast barium enema or fecal occult blood testing. Colonoscopy is the only screening examination accepted at 10-year intervals among people at average risk without significant polyps found. It allows direct visualization of the entire colon and rectum and permits removal of adenomatous polyps, precursors of colorectal cancer. The frequency of polypectomy is an important indicator of quality assurance for colorectal cancer screening programs.

In the province of Quebec, physicians are reimbursed for medical services by the Régie de l'Assurance Maladie du Québec (RAMQ), the government agency responsible for administering the provincial health insurance plan. Physicians receive additional remuneration for performing a polypectomy if they include the procedure code in their claim.

Data from physician claims databases are commonly used in health services research,^{1–7} even though the data are collected for administrative purposes and physician reimbursement.

Procedure codes in physician claims databases are presumed to have a very high level of agreement with data in medical charts.⁸ A physician making a claim will need to submit the diagnostic code and, when applicable, the procedure code. Studies that rely on physician claims databases can be divided into those that examine the diagnostic codes entered and those that examine the procedure codes entered. Few studies have attempted to validate procedure codes, and often not as the primary study objective.^{9–14}

We conducted a study to determine the level of agreement between physician claims for polypectomy and documentation of the procedure in endoscopy reports.

Methods

Study design and population

This retrospective study stems from a larger prospective cohort study aimed at developing an algorithm to identify screening colonoscopies in physician claims databases in three provinces. For our study, we used data from the Quebec database (RAMQ). The study sites were in Montréal, Que., and included the Montréal General Hospital, the

Royal Victoria Hospital, St. Mary's Hospital Centre, the Jewish General Hospital, Hôtel Dieu, Hôpital Maisonneuve-Rosemont and Hôpital Fleury. Six are teaching hospitals (four affiliated with McGill University, and two with Université de Montréal), and one is a community hospital.

Participants were staff endoscopists whose patients underwent colonoscopy between January and March 2007. Endoscopists were eligible if they performed colonoscopies and were remunerated by the RAMQ. On days when the research assistant attended, consecutive patients of participating endoscopists were approached. Patients were included if they were 50–80 years old and underwent scheduled colonoscopy (whether or not the entire colon was visualized). Those who were not eligible for coverage under the provincial health insurance plan in the prior year or were unable to provide informed consent were excluded.

Data collection

Polypectomy data were collected from both the endoscopy report (medical chart) and the RAMQ database. Endoscopy reports were reviewed by two research assistants blinded to the status of the polypectomy procedure code in the database. Both data abstractors had medical training, and one was a clinical gastroenterologist. Information on whether or not polypectomy was performed was obtained using a standardized data abstraction form, which also collected the patient's identification number, the hospital site and the date of the colonoscopy. Records from the RAMQ database linked to the patient's identification number were obtained for the date of the index colonoscopy and for the two months before and after the procedure. The presence of the polypectomy procedure code (0749) in the database record for the index colonoscopy was noted.

Ethics approval

Ethics approval was obtained from the Institutional Review Board of the McGill University Faculty of Medicine. Written informed consent was obtained from endoscopists and patients. Only identification numbers for participating patients and endoscopists appeared on forms.

Statistical analysis

The documentation of any polypectomy performed during the index colonoscopy in the endoscopy report was compared with the presence of the procedure code for polypectomy during the index colonoscopy in the RAMQ database. Two-by-two tables were compiled and measures of accuracy calculated, including sensitivity, specificity, positive predictive value, negative predictive value,

concordance and kappa value.^{15,16} Concordance was defined as the degree of interrater agreement between the presence of the procedure code in the database and documentation of polypectomy in the endoscopy report. Discordance was defined as the sum of false-positive and false-negative results divided by the total number of colonoscopies. The kappa value indicated the probability of agreement after adjusting for agreement by chance.¹⁷ All parameter estimates are presented, including 95% confidence intervals (CIs) calculated from the binomial distribution using R version 2.7.0.

To account for possible correlations arising from patients being nested with physician practices, and physicians being nested within hospital sites, we fit a Bayesian hierarchical model using the statistical program WinBUGS version 1.4.3 (www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). The first level used a binomial model to estimate the probability of outcomes for each physician. A logit-transformation was then applied to the binomial parameters (one from each physician) and was subject to a second hierarchical level to account for hospitals. Noninformative prior distributions were used.

The original prospective cohort study enrolled patients across three provinces. Assuming the endoscopy report as the "gold standard," the desired sample size was 500 patients from each province. This implied accuracy of a 95% CI to ± 0.046 for estimating sensitivity, and ± 0.029 for specificity.

Results

During the study period, 707 patients were enrolled and underwent colonoscopy by 38 endoscopists. We excluded three endoscopists (accounting for 10, 4 and 4 colonoscopies respectively) who could not submit RAMQ claims because of their fee schedule, which left 35 endoscopists and 689 patients. Polypectomy data were abstracted for all 689 patients. Instances of missing data were included as part of the validation for the polypectomy procedure code.

Characteristics of the patients and endoscopists are summarized in Table 1. Thirty-one endoscopists (88.6%) were gastroenterologists. On average, each endoscopist performed 19.7 (standard deviation [SD] 16.5) colonoscopies. The patients had a mean age of 60 (SD 7.0) years, and 49.3% were female. Screening for colon cancer was the indication for 344 (49.9%) of the colonoscopies. Completion rates for colonoscopy (determined from physician claims for procedure code 0697) were 98.0% for screening and 94.5% for nonscreening colonoscopies. The mean number of colonoscopies per hospital site was 98.4 (SD 54.2).

Colonoscopy procedure codes

Using the date of the index colonoscopy \pm one day, we noted that a colonoscopy procedure code was entered in the RAMQ database for 667 (96.8%) of the 689 patients. Using the date of the index colonoscopy \pm two months, we identified eight additional patients as having had a colonoscopy (mean 6.9 [range 2–15] days from the index colonoscopy). The remaining 14 colonoscopies (2.0%) in our study that were not identified in the database were performed by 12 endoscopists (mean 1.2, range 1–2) at six of the seven study sites. In two instances, a procedure code for polypectomy was submitted without a procedure code for colonoscopy.

Accuracy of claims for polypectomy

Table 2 shows the level of agreement between the physician claims for polypectomy in the RAMQ database and the documentation of the procedure in the endoscopy reports. The sensitivity of physician claims for polypectomy in the database was 84.7% (95% CI 78.6%–89.4%), and the specificity was 99.0% (95% CI 97.5%–99.6%). The positive predictive value was 97.0% (95% CI 92.7%–

98.9%), and the negative predictive value was 94.5% (95% CI 92.0%–96.2%). Concordance was 95.1% (95% CI 93.1%–96.5%). The kappa value was 0.87 (95% CI 0.83–0.91), which indicated a very high probability of agreement after adjustment for agreement by chance.

Overall, 29 (15.3%) of the 190 polypectomies noted in the endoscopy reports were not identified in the database (false-negative result). Conversely, the database identified five polypectomies (0.7%) that were not recorded in the endoscopy reports (false-positive result). In addition, we noted that 2 of the 161 claims for polypectomy in the database were submitted without a colonoscopy procedure code (identified by expanding the database search to include two months before and after the index colonoscopy). Using only the index colonoscopy date, we found that the sensitivity decreased to 83.7% (159/190; 95% CI 77.5%–88.5%).

For the 35 endoscopists, the level of discordance was low (4.9% [34/689], 95% CI 3.3%–6.6%); no meaningful clinical differences were found between the endoscopists. Similarly, there were no meaningful clinical differences in the level of discordance between the study sites.

The Bayesian hierarchical model had measures of accuracy similar to those in the binomial analysis. The sensitivity was 90.2% (95% credible interval 57.5%–99.1%), the specificity was 99.9% (95% credible interval 98.9%–100.0%), the positive predictive value was 99.2% (95% credible interval 95.0%–100.0%), the negative predictive value was 96.0% (95% credible interval 88.6%–99.2%) and concordance was 96.0% (95% credible interval 90.4%–98.8%). One exception was the lower limit of the credible interval for sensitivity, which was much lower than the lower limit of the confidence interval in the binomial analysis.

Interpretation

We found a high level of agreement between the procedure code for polypectomy in the RAMQ claims database and data contained in the endoscopy report. Concordance was high (95.1%), as

Table 1: Characteristics of patients and endoscopists included in the study	
Characteristic	Value
Patients (<i>n</i> = 689)	
Age, yr, mean (SD)	60 (7.0)
Female sex, no. (%)	340 (49.3)
Endoscopists (<i>n</i> = 35)	
No. of colonoscopies per endoscopist, mean (SD)	19.7 (16.5)
No. of colonoscopies per site, mean (SD)	98.4 (54.2)
Specialty, no. (%)	
Gastroenterologist	31 (88.6)
Other	4 (11.4)
Complete colonoscopies, no. (%)	
Screening	337/344 (98.0)
Nonscreening	326/345 (94.5)
No. of endoscopists [colonoscopies] per site	
A	5 [94]
B	7 [118]
C	4 [177]
D	8 [142]
E	5 [94]
F	5 [44]
G	2 [20]
Note: SD = standard deviation.	

Table 2: Level of agreement between physician claims for polypectomy and documentation of procedure in endoscopy reports for 689 colonoscopies

Polypectomy code in physician claim	Polypectomy noted in endoscopy report		Total
	Yes	No	
Yes	161	5	166
No	29	494	523
Total	190	499	689

was the kappa value (0.87). However, the sensitivity was lower (84.7%). It was also a conservative estimate, given that two of the claims for polypectomy had the incorrect date. If those procedures were excluded, the sensitivity would be decreased further, to 83.7%. The Bayesian hierarchical model had measures of accuracy similar to those of the binomial analysis, except that the lower bound of the credible interval for sensitivity was much lower than the lower limit of the confidence interval in the binomial analysis. This difference occurred because of variability between participating endoscopists and between study sites.

Random administrative error is our primary hypothesis for the lower sensitivity. Physicians do not need to include a polypectomy code when they submit a claim for colonoscopy. There are many opportunities for random administrative error to occur from the time of polypectomy to the procedure code appearing in the RAMQ database. Some physicians use online billing programs, whereas others use third-party billing agents. An omission translates to less remuneration and is unlikely to occur purposefully.

When we expanded our search of the RAMQ database by two months before and after the index colonoscopy to capture claims submitted with an incorrect date, we found an additional eight colonoscopies. These eight procedures had been entered in the database within 14 days after the true index date. We believe these colonoscopies represent administrative errors, since it would be highly unusual to perform two elective colonoscopies within two weeks in the same patient.

Procedure codes for gastroscopy, colonoscopy, flexible sigmoidoscopy and polypectomy have been used previously to assess population-based procedure rates and resource utilization in gastroenterology in Canada.^{18–22} However, none of the studies validated the procedure codes. Only two studies (both in the United States) provided information on the accuracy of procedure codes in gastroenterology. One examined codes for colonoscopy and flexible sigmoidoscopy compared with medical records of the primary care physician (not the physician submitting the claim).²³ With the medical record as the “gold standard” for colonoscopy, the authors reported that the procedure codes in the Medicare claims database had a sensitivity of 94%, a specificity of 95%, a positive predictive value of 91%, a negative predictive value of 97%, concordance of 95% and a kappa value of 0.89. The other study examined procedure codes in the national Veterans Health Administration database and found a specificity (the only measure of accuracy reported) for colonoscopy of more than 99%.²⁴

Strengths and limitations

Strengths of our study include the following: the colonoscopies performed represented real-world practices for endoscopists, since patient lists were unaltered and the study involved multiple endoscopists at multiple sites; RAMQ is a physician claims system that operates with the same diagnostic and procedure codes throughout the province; and physicians were unaware of the purpose of our study, so submitted claims were unlikely to be biased.

Three study limitations are noteworthy. First, the generalizability of our findings is limited. All but 20 colonoscopies were performed in academic centres in Montréal. Physicians from different academic centres are likely to have widely variable billing practices. For example, in the first author’s academic centre, claims are submitted several ways: by the gastroenterologists (either by completing forms by hand or using an online billing program) or by secretaries or other billing agents via an online billing program. Given individual physician characteristics, different strategies for submitting claims within one hospital, and the fact that there was no detected variability in accuracy between physicians and between the study sites, we believe that our results are representative of other Quebec institutions, including perhaps private free-standing facilities. However, we were unable to assess the representativeness of our sample. The optimal study design would be to select random samples from among all Quebec patients who underwent colonoscopy at various sites, identified through hospital records or the RAMQ database primarily, and compare the administrative data with information documented in the endoscopy reports.

Second, there may have been selection bias, because participating endoscopists may have been more (or less) meticulous in their recording of information in the medical charts or payment claims.

Third, there may have been misclassification owing to the difficulty in defining polypectomy in the endoscopy report, since some endoscopy reports did not specify the mechanism of polyp removal. For example, a small polyp removed using only a biopsy forceps (and not a snare) should be recorded as a polypectomy and submitted as procedure code 0749. If some of these polyps were submitted as a biopsy (procedure code 0750), it may explain in part the lower sensitivity. We did not collect data for code 0750 from the database. However, remuneration is higher for polypectomy than for biopsy, so this potential coding misclassification would likely have been uncommon. Removal of a polyp of any size is considered a polypectomy; there is no

size requirement. Physicians are relied on ultimately to submit claims that accurately represent the procedure performed.

Another “gold standard” against which the administrative data could have been compared is the pathology report. However, we felt that use of the endoscopy report was superior because (a) a polyp can be removed without tissue being recovered, and thus no pathology report would exist; and (b) if polypectomy was not performed, there would be no pathology report, and one would need the endoscopy report to determine whether a pathology report was missing. In addition, the endoscopy report is completed immediately after the procedure, which minimizes recall bias. In our study all endoscopy reports were found.

Conclusion

Despite providing a reasonably accurate estimate of the frequency of polypectomy, physician claims underestimated the number of procedures by more than 15%. Such differences could affect conclusions regarding quality assurance if used to evaluate population-based colorectal cancer screening programs. Procedure codes are assumed to be accurate in physician claims databases, and few investigators routinely validate these data. Our findings reinforce the need for investigators of future studies to consider validating codes before interpreting their findings, especially when health care policy may be influenced by the results. The present study lays the groundwork for conducting further research in resource allocation and quality control for colorectal cancer screening in Quebec and elsewhere in Canada.

References

- Beitel AJ, Olson KL, Reis BY, et al. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 2004; 20:355-60.
- Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 2000;53:1258-67.
- Maselli JH, Gonzales R. Measuring antibiotic prescribing practices among ambulatory physicians: accuracy of administrative claims data. *J Clin Epidemiol* 2001;54:196-201.
- McKnight J, Scott A, Menzies D, et al. A cohort study showed that health insurance databases were accurate to distinguish chronic obstructive pulmonary disease from asthma and classify disease severity. *J Clin Epidemiol* 2005;58:206-8.
- Tamblyn R, Reid T, Mayo N, et al. Using medical services claims to assess injuries in the elderly: sensitivity of diagnostic and procedure codes for injury ascertainment. *J Clin Epidemiol* 2000;53:183-94.
- Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004; 57:131-41.
- Muhajarine N, Mustard C, Roos LL, et al. Comparison of survey and physician claims data for detecting hypertension. *J Clin Epidemiol* 1997;50:711-8.
- Mitchell JB, Ballard DJ, Whisnant JP, et al. Using physician claims to identify postoperative complications of carotid endarterectomy. *Health Serv Res* 1996;31:141-52.
- Qureshi AI, Harris-Lane P, Siddiqui F, et al. International classification of diseases and current procedural terminology codes underestimated thrombolytic use for ischemic stroke. *J Clin Epidemiol* 2006;59:856-8.
- Javitt JC, McBean AM, Sastry SS, et al. Accuracy of coding in Medicare part B claims. Cataract as a case study. *Arch Ophthalmol* 1993;111:605-7.
- Quan H, Parsons GA, Ghali WA. Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. *Med Care* 2004;42:801-9.
- Duszak R, Blackham WC, Kusiak GM, et al. CPT coding by interventional radiologists: a multi-institutional evaluation of accuracy and its economic implications. *J Am Coll Radiol* 2004;1: 734-40.
- Abraham NS, Cohen DC, Rivers B, et al. Validation of administrative data used for the diagnosis of upper gastrointestinal events following nonsteroidal anti-inflammatory drug prescription. *Aliment Pharmacol Ther* 2006;24:299-306.
- Ladouceur M, Rahme E, Pineau CA, et al. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics* 2007;63:272-9.
- Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. *J Natl Cancer Inst* 1993;85:566-70.
- Hiatt RA, Perez-Stable EJ, Quesenberry C Jr, et al. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Prev Med* 1995;24:278-85.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Bressler B, Paszat LF, Vinden C, et al. Colonoscopic miss rates for right-sided colon cancer: a population-based analysis. *Gastroenterology* 2004;127:452-6.
- Hilsden RJ. Patterns of use of flexible sigmoidoscopy, colonoscopy and gastroscopy: a population-based study in a Canadian province. *Can J Gastroenterol* 2004;18:213-9.
- LeLorier J, Page V, Castilloux AM, et al. Management of new symptoms of dyspepsia in the elderly in Quebec. *Can J Gastroenterol* 1997;11:669-72.
- Rabeneck L, Paszat LF. A population-based estimate of the extent of colorectal cancer screening in Ontario. *Am J Gastroenterol* 2004;99:1141-4.
- Schultz SE, Vinden C, Rabeneck L. Colonoscopy and flexible sigmoidoscopy practice patterns in Ontario: a population-based study. *Can J Gastroenterol* 2007;21:431-4.
- Schenck AP, Klabunde CN, Warren JL, et al. Data sources for measuring colorectal endoscopy use among Medicare enrollees. *Cancer Epidemiol Biomarkers Prev* 2007;16:2118-27.
- Fisher DA, Gruber JM, Castor JM, et al. Ascertainment of colonoscopy indication using administrative data. *Dig Dis Sci* 2010; 55:1721-5.

Affiliations: From the Division of Gastroenterology (Wyse), Department of Medicine, Jewish General Hospital, McGill University, Montréal, Que.; the Department of Epidemiology and Biostatistics (Joseph), McGill University, Montréal, Que.; the Divisions of Gastroenterology (Barkun) and Clinical Epidemiology (Barkun, Sewitch) and the Department of Medicine (Barkun, Sewitch), McGill University Health Centre, Montréal, Que.

Contributors: Jonathan Wyse and Maida Sewitch were responsible for study conception and design. Jonathan Wyse was responsible for data collection and drafting of the article. Maida Sewitch was responsible for providing the database of patients for the study derived from the original prospective study. Lawrence Joseph provided statistical expertise. All of the authors were responsible for analysis and interpretation of the data, critical revision of the article for important intellectual content and final approval of the article submitted for publication.

Funding: This research was funded by the Canadian Cancer Society (grant no. 017054) through an operating grant awarded to Maida Sewitch. Jonathan Wyse was supported by a bursary from the Research Institute of McGill University Health Centre. Lawrence Joseph is a Chercheur National of the Fonds de la recherche en santé du Québec (FRSQ). Alan Barkun is a Chercheur National of the FRSQ and holder of the Douglas G. Kinnear Chair in Gastroenterology at McGill University. Maida Sewitch is supported as a Research Scientist of the Canadian Cancer Society through an award from the National Cancer Institute of Canada.