ORIGINAL RESEARCH

Validation of the Diagnostic Algorithms for 5 Chronic Conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): A Kingston Practice-based Research Network (PBRN) Report

Amjed Kadhim-Saleh, BHSc, MSc, Michael Green, MD, MPH, Tyler Williamson, PhD, Duncan Hunter, PhD, and Richard Birtwhistle, MD, MSc

Objective: **The objective of this study was to assess the validity of electronic medical records–based diagnostic algorithms for 5 chronic conditions.**

Methods: **A retrospective validation study using primary chart abstraction. A standardized abstraction form was developed to ascertain diagnoses of diabetes, hypertension, osteoarthritis, chronic obstructive pulmonary disease, and depression. Information about billing, laboratory tests, notes, specialist and hospital reports, and physiologic data was collected. An age-stratified random sample of 350 patient charts was selected from Kingston, Ontario, Canada. Approximately 90% of those charts were allocated to people aged ≥60 years.**

Results: **Three hundred thirteen patient records were included in the study. Patients' mean age was 68 years and 52% were women. High interrater reliability was indicated by 92% complete agreement and a κ statistic of 89.3%. The sensitivities of algorithms were 100% (diabetes), 83% (hypertension), 45% (osteoarthritis), 41% (chronic obstructive pulmonary disease), and 39% (depression). The lowest specificity was 97%, for depression. The positive predictive value ranged from 79% (depression) to 100%, and the negative predictive value ranged from 68% (osteoarthritis) to 100%.**

Conclusions: **The diagnostic algorithms for diabetes and hypertension demonstrate adequate accuracy, thus allowing their use for research and policy-making purposes. The algorithms for the other 3 conditions require further refinement to attain better sensitivities. (J Am Board Fam Med 2013;26: 159–167.)**

Keywords: **Chronic Disease, Health Informatics, Practice-based Research**

Chronic diseases constitute a major burden of illness in Canada and around the world. Recent estimates suggest that 46% of adult Canadians suffer from one or more of 7 common chronic diseases.[1] Of these conditions, 6 million Canadians are affected with hypertension,[2] 2 million with diabetes,[3] 1.2 million with major depression,[4] >750,000 adults with chronic obstructive pulmonary disease (COPD),[5] and 3 million with osteoarthritis.[6]

Currently available information on chronic diseases at the national level is derived from databases such as hospital discharge summaries, disease-specific registries, and population health surveys. These sources have significant limitations, such as the inability to capture data on conditions that do not lead to hospitalizations and the unreliability of self-reported surveys.[7] A large validation study of the Discharge Abstract Database concluded that coding of comor-

This article was externally peer reviewed.
Submitted 16 July 2012; revised 25 September 2012; accepted 16 October 2012.

From the Departments of Community Health and Epidemiology (AK-S, MG, DH) and Family Medicine (MG, DH, TW, RB), Queen's University, Kingston, Ontario, Canada.

Funding: The CPCSSN project is funded by the Public Health Agency of Canada (PHAC). The views expressed in this manuscript do not necessarily represent the views of PHAC.

Conflict of interest: none declared.

Corresponding author: Richard Birtwhistle, Department of Family Medicine, Queen's University, 220 Bagot St., Kingston, ON K7L 5E9 Canada (E-mail: birtwhis@queensu.ca).

bidities was poor.[8] For example, the complete agreement between the original record and the re-abstracted record was only 3.7% for COPD before admission and 8% for diabetes.[8] Surveys also are limited in their use for ongoing surveillance because of added financial burden.[9]

At the Canadian provincial level, billing for physician services may provide a source of data, but it is limited in the depth of information because administrative data are created for financial management rather than research purposes.[10] When compared against a clinical research database, administrative data had only 20% agreement.[11]

Primary care databases constitute another source of data on chronic conditions. For instance, people with one or more chronic conditions accounted for 51% of family physician encounters,[12] suggesting that comprehensive clinical records collected by primary care physicians could be a rich resource for researchers and policymakers. The benefit of using primary care databases is that they provide prospective and systematic collection of clinically verified data that can be comprehensive for studying a variety of important outcomes.[13]

The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) is one example of a primary care database. It is an electronic medical records (EMR)–based information system for chronic disease surveillance that has been functioning since April 2008. It brings together sentinel practices in 10 practice-based research networks (PBRNs) across the country and institutional partners including academic research centers and departments, the College of Family Physicians of Canada, and the Canadian Institute for Health Information.[7] Longitudinal data are extracted from the participating practices every 3 months and include the following information: network and provider identifiers, de-identified patient demographics, encounter date and type, health conditions, risk factors, referrals, laboratory investigations, procedures, and medications. There are currently >300,000 patient records included in the database, with 2 to 3 years of data extraction already available.

CPCSSN relies on diagnostic algorithms to identify patients with chronic conditions. Diagnostic algorithms are protocols that use various indicators, such as billing data, laboratory test results, and medications, to ascertain diagnoses in the database.

To use a primary care database, it is important to investigate the quality of its data. There are a number of factors that contribute to diagnostic inaccuracy and incompleteness, including misclassification, missing data, lack of standardization, and data usability limitations, thus highlighting the need for a validation study that quantitatively assesses diagnostic accuracy.

Misclassification can occur because of difficulty in differentiating between complex conditions, underreporting of conditions with more subjective criteria, misclassifying a tentative diagnosis as definitive, as well as excluding less severe cases that do not require extensive treatment.[14–16] Data could be missing when use of external health care is not recorded, thus limiting the completeness of computer-based diagnostic algorithms.[10] Lack of standardization may occur when different EMR system platforms are used. Even within the same EMR system, the level of accuracy and completeness may vary across different providers and sites.[8,15,17] Finally, there are a number of limitations to data usability, such as the inability to code certain fields (eg, physician notes).[10,18]

Many validation studies have been conducted to evaluate the accuracy of computer-based diagnostic algorithms. Herrett et al[16] conducted a systematic review of validation studies in the United Kingdom-Clinical Practice Research Datalink. A total of 357 validations on 183 diagnoses were reviewed. Although estimates of validity were generally high, the review highlighted a number of serious limitations. First, the quality of reporting was insufficient to assess bias and generalizability across the database, thus hindering the interpretation of findings. For example, many studies did not provide the medical codes that were used to define the index diseases. In addition, there was insufficient detail in the sampling strategy, the percentage of missing data, and whether reviewers were blinded to diagnosis. Second, the majority of validation studies were conducted on a highly selected cohort of patients that considered only cases rather than selecting a random sample of both cases and noncases. This case-stratified sampling led to an inability to estimate sensitivity and specificity, and thus only the positive predictive value (PPV) was calculated. Although informative, the PPV estimate was dependent on the prevalence of the condition, unlike sensitivity and specificity measures. Finally, the response rate for many of the validation studies was generally low, ranging 55% to 100%.

The aim of this study is to conduct a validation of EMR-based diagnostic algorithms that addresses these limitations by specifying the details of data to be reported using a random sample that will validate both cases and noncases and having full participation. Improving the understanding of the quality of data in EMR-based information systems such as CPCSSN will allow for the identification of areas where the data can be considered reliable and useful for research and decision making by understanding the limitations of data acquired in this manner and improving the quality of such data in the future. The objective of this study was to assess the validity of EMR-based diagnostic algorithms for 5 chronic conditions in the CPCSSN database.

Methods

Design

This validation study of case-finding diagnostic algorithms was a retrospective analysis of EMR-based primary care data. Figure 1 illustrates a comparison of the level of concordance between 2 data sources: the CPCSSN database and the primary abstraction of patients' EMRs. This was a pilot study for the larger CPCSSN validation project, which will take place within the CPCSSN. Age-stratified sampling was used to ensure that the prevalence of the index conditions, especially COPD, would be sufficiently high. This, in turn, would ensure that the width of the 95% confidence interval would be maintained at ≤10%, an acceptable level of precision.

Study Population

All patients who attended the Kingston (Ontario) PBRN and all practices within the network (n = 22) were included in the study. The Kingston PBRN is one of the 10 networks within CPCSSN, and EMRs had been implemented at this site since 2004. An age-stratified random sample of 350 patient charts was selected from the entire patient pool, with 90% of these charts allocated to people aged 60 years or older.

Definition of Variables and Data Collection

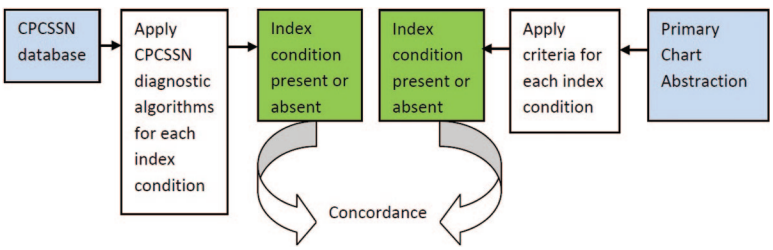
CPCSSN Diagnostic Algorithms

Case-finding diagnostic algorithms have been developed to identify patients with chronic conditions. Because of space limitations, we did not include the algorithms in this article; however, the CPCSSN diagnostic algorithms have been published previously.[7] These algorithms are based on various indicators, including billing data, laboratory test results, and medications to ascertain diagnoses. For example, diabetes can be identified from existence of billing data (code 250.X), medications (insulin, glyburide, metformin), and laboratory tests (hemoglobin A1C >0.07, fasting blood sugar >7 mmol/litre). The *International Statistical Classification of Diseases and Related Health Problems* (version 9) was used along with certain drugs and positive test results to ascertain diagnosis.[7] The application of CPCSSN algorithms was conducted after the completion of the primary chart abstraction to ensure that abstractors were blinded to the CPCSSN diagnosis.

Primary Chart Abstraction

The gold standard in this study was the primary audit of electronic records undertaken by 3 trained and experienced researchers (AJK). EMRs were reviewed electronically to determine whether pa-

Figure 1. A visual representation of the study design. Computerized data collected through the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) will be compared against the gold standard (primary chart abstraction). For each index condition, a CPCSSN protocol algorithm was used to determine whether a patient had the disease. In contrast, prespecified criteria were used in chart abstraction to identify the presence or absence of index conditions. The level of concordance will be compared between the 2 data sources.



tients had any of the index conditions, along with the location of information used to make this assessment. A standardized abstraction form was developed based on consultations with clinicians who used the EMR platform at the Kingston PBRN. Information on patient's age and sex, health conditions, medications, physiologic data (weight, height, body mass index, and blood pressure readings), test results, referrals, procedures, hospitalizations, billing data, physician notes, and specialist and hospital reports were collected. Medical record abstractors used Microsoft Access (Microsoft Corp., Redmond, WA) to implement the standardized abstraction form and record the data. Abstractors had to examine the entire record to reach a conclusion on whether the patient had the index condition. It is also important to emphasize that abstractors were blinded to the CPCSSN diagnoses. The first 10 EMRs reviewed by research associates were re-abstracted independently by a second associate to ensure adequate quality and consistency of the data. Reviewers then consulted with a clinician (MG, RB), who arbitrated cases of disagreement.

Patients were classified into 1 of 3 categories: (1) confirmed positive; (2) confirmed negative; (3) unsure/untested. For example, a patient was classified as confirmed positive for diabetes if they were prescribed metformin regularly and the physician notes confirmed that the patient had diabetes. A patient with none of the indicators for diabetes and a normal level of hemoglobin A1C was classified as a confirmed negative. Patients were classified as unsure/untested when there were neither indicators for the disease nor any tests done to rule out the condition. Patients also were classified as unsure when there was contradictory evidence. In those cases, consultation with a clinician (MG) was sought to reach a decision.

Analysis

All calculations were done using SAS version 9.2 (SAS, Inc., Cary, NC). Sensitivity, specificity, PPVs, and negative predictive values (NPVs) were provided, along with 95% confidence intervals. Two-by-two contingency tables are provided for each index condition, specifying whether a condition is present or absent according to CPCSSN diagnostic-algorithms versus primary chart abstraction. The level of interrater reliability was expressed as a kappa statistic that takes into consideration agreement that occurs by chance as well as percent agreement.[19]

Sensitivity was calculated by dividing the number of records in which a diagnosis was present according to both sources (CPCSSN and EMR abstraction) by the total number of true cases. Specificity was calculated by dividing the number of records in which a diagnosis is absent from both sources (CPCSSN and EMR abstraction) by the total number of true noncases. When calculating sensitivity and specificity, patients classified as confirmed negative or unsure/untested were considered to lack the index disease according to record abstraction. To account for the inherent clustering in primary care data within physicians, confidence intervals were calculated through inflating the variance by the appropriate design effect using the estimated intraclass correlation. The study protocol was approved by the Queen's Health Sciences Research Ethics Board.

Results

The study population comprised 313 patients. Thirty-seven patient records were excluded from the analysis because of a lack of data (nonrostered patient or movement to a nursing home [n = 18], patient death [n = 15], and patient left practice [n = 4]). The mean age of patients was 68 years, and 52% were women. In a sample of 10 that were reviewed independently, the exact percentages of agreement between reviewers were 100% (diabetes, osteoarthritis), 90% (COPD, depression), and 80% (hypertension). Thus, the overall complete agreement between record abstractors was 92%. The overall kappa statistic was 89.3%. In approximately 5% of patient records, a clinician (MG) was consulted on the appropriate diagnoses. The clinician examined the full record to reach a decision.

Based on medical record abstraction, approximately 80% of patients had at least 1 of the 5 chronic conditions. The prevalence of comorbidities was 31% for patients with hypertension and osteoarthritis, 16% for patients with hypertension and diabetes, and 13% for patients with hypertension and depression. Approximately 8% of patients had hypertension, diabetes, and osteoarthritis. Table 1 provides further details on the sample demographics.

Table 2 provides details on the concordance and discordance between the 2 data sources: EMR ab-

Table 1. Demographics of Patients Included in The Study Sample from the Kingston Practice-Based Research Network

Demographics	Count (n = 313)
Male	150 (48)
Female	163 (52)
Age, yr	
≥60	282 (90)
<60	31 (10)
Chronic conditions	
Diabetes	53 (17)
Hypertension	181 (58)
Osteoarthritis	144 (46)
COPD	29 (9)
Depression	66 (21)
Number of chronic conditions	
At least 1 of 5	250 (80)
2	94 (30)
3	37 (12)
4	17 (5)

Values provided as n (%).
COPD, chronic obstructive pulmonary disease.

straction and CPCSSN algorithms. It also provides details on the sensitivity, specificity, PPV, and NPV, along with the exact 95% confidence intervals. Using an intraclass correlation of 0.035 and an average cluster size of 14 patients per physician, the estimated design effect was 1.455. Therefore, all confidence intervals were 1.206 (square root of 1.455) times wider than if there were no cluster effect.

Based on abstraction, 16.9% of patients had type 2 diabetes. The diagnostic algorithm for identifying patients with diabetes had 100% sensitivity, 99% specificity, 95% PPV, and 100% NPV. Hypertension was diagnosed in 57.8% of patients. The diagnostic algorithm for identifying patients with hypertension had 83% sensitivity, 98% specificity, 98% PPV, and 81% NPV. Osteoarthritis was diagnosed in 46% of patients. The diagnostic algorithm for identifying patients with osteoarthritis had 45% sensitivity, 100% specificity, 100% PPV, and 68% NPV. COPD was diagnosed in 9.2% of patients. The diagnostic algorithm for identifying patients with COPD had 41% sensitivity, 99% specificity, 80% PPV, and 94% NPV. Finally, depression was diagnosed in 21% of patients. The diagnostic algorithm for identifying patients with depression had 39% sensitivity, 97% specificity, 79% PPV, and 86% NPV.

Discussion

Summary of Main Findings

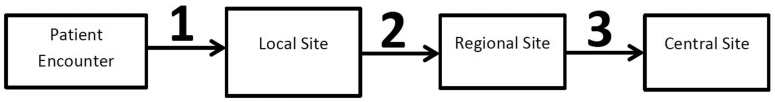
The CPCSSN project was initiated to provide ongoing surveillance of chronic conditions. In addition, the CPCSSN data could be used by researchers to identify a cohort of patients and examine the effectiveness of prevention and management strategies. Policymakers could also use this data to plan and allocate resources needed to manage chronic conditions. The findings suggest that the specificity for all 5 conditions was very high (lowest for depression [97%]). Thus, the diagnostic algorithms are highly specific and yield few false-positive cases. Sensitivities of the CPCSSN algorithms, in contrast, varied considerably among the 5 conditions. The sensitivity of diagnostic algorithms for diabetes (100%) and hypertension (83%) was adequate. Thus, the majority of true cases of diabetes and hypertension are being identified correctly by the CPCSSN algorithms. However, the sensitivities for the other algorithms (osteoarthritis, COPD, depression) were significantly low. This suggests that the current algorithms used are underestimating the true prevalence of these 3 conditions.

Table 2. Validation Results of the Case-Finding Diagnostic Algorithm for the 5 Chronic Conditions

Condition	Chart+ CPCSSN+	Chart- CPCSSN+	Chart+ CPCSSN-	Chart- CPCSSN-	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Diabetes	53	3	0	257	100% (92–100)	99% (96–100)	95% (83–100)	100% (98–100)
Hypertension	150	3	31	129	83% (75–89)	98% (93–100)	98% (94–100)	81% (72–88)
Osteoarthritis	65	0	79	169	45% (35–55)	100% (97–100)	100% (93–100)	68% (61–75)
COPD	12	3	17	281	41% (20–65)	99% (97–100)	80% (46–99)	94% (90–97)
Depression	26	7	40	240	39% (25–55)	97% (94–99)	79% (57–94)	86% (80–90)

CI, confidence interval; COPD, chronic obstructive pulmonary disease; CPCSSN, Canadian Primary Care Sentinel Surveillance Network; NPV, negative predictive value; PPV, positive predictive value.

Figure 2. Data flow from physician's office to Canadian Primary Care Sentinel Surveillance Network's (CPCSSN's) central repository site. At stage 1, the data are entered into the electronic medical record. At this stage, there is a possibility of nondocumentation that can result either from the patient not reporting the condition (or being unaware of it) or physician failure to document the condition in the chart. At stage 2, CPCSSN extracts data from the local sites. At this stage, certain data, including physician notes, hospital and specialist reports, imaging reports, past medical and social history, certain questionnaires (eg, depression screening), procedures, and imaging reports, are not extracted by CPCSSN. These data were only available through the manual chart abstraction. At stage 2 patient identifying information is removed to ensure privacy. At stage 3 cleaned and transformed data is uploaded to the central database.



Strengths and Limitations

This study is the first validation of the CPCSSN case-finding diagnostic algorithms. Since CPCSSN providers have already agreed to allow audits of medical records, there was no issue in accessing data. This demonstrates a significant improvement over many previous validation studies that had low response rates. For example, a validation study using an administrative database in Ontario had a response rate of only 11%.[9] A low response rate may have led to selection bias and a lack of generalizability because responding practices are systematically different from nonresponding practices.[16] In this study, medical records from all randomly selected practices were examined to assess the presence or absence of index conditions.

The sample size was relatively large compared with similar studies done in the U.K. Clinical Practice Research Datalink.[14] Also, the random sampling technique used in this study improves on previous research that employed disease-specific stratified sampling, which is based on information for cases alone. This type of sampling is limited in that it cannot be used to calculate sensitivity and specificity.[16] The PPV is dependent on the prevalence of disease, and thus it might not provide a complete picture of the validity of diagnostic algorithms, even if it was found to be high.[16] This study addresses this limitation by employing an age-stratified random sample from all patients. By validating noncases, this study ensured that all patients (both cases and noncases) were subject to the same criteria.[16] It also avoided verification bias, which results from assessing the accuracy of a test on cases only.[20]

This study has important limitations. One challenge was the use of different EMRs by research networks. All EMRs have different coding structures; data extraction was also different.[7] Since data from only one EMR platform was analyzed, it was not possible to assess whether regional differences and different EMRs have an impact on measures of validity. This study was a pilot project for the larger CPCSSN validation. At least 2 other similar projects are underway at 2 sites in Alberta and Manitoba. Once these findings are known, validity measures can be compared across the different sites.

Figure 2 details the flow of data from the physician's office to CPCSSN's central repository site, indicating areas where data are lost. One limitation to consider is the potential for nondocumentation in patient records. This issue can arise from either the patient not reporting the condition or the provider failing to document in the records. In both cases, assuming that the patient does not have the condition overestimates the accuracy of the diagnostic algorithms. However, this issue of nondocumentation is mitigated by the thoroughness of the manual abstraction and by relying on not only physician notes, but also laboratory test results, medications, hospital records, and billing data. The physician would be unlikely to omit documentation for a condition for which he or she was prescribing a medication or submitting a bill.

The study could not assess the generalizability of findings across the CPCSSN database because it examined accuracy at the Kingston PBRN alone. In other words, data from one network may not represent the overall data quality. When calculating sensitivity and specificity, patients who were classified as untested/unsure were considered "presumed negative" because they had none of the indicators for the conditions. Therefore, patients who were

confirmed and presumed negative were combined as lacking the disease. It is possible that some of the patients presumed negative had the condition; however, this is unlikely because these patients had none of the various indicators (eg, medications, test results, physician notes, disease registries). It is also important to acknowledge the limitation that the study is subject to measurement error since a true gold standard is not available. Medical record abstraction is an imperfect gold standard and thus is subject to measurement error.[20] However, the use of primary record abstraction is considered the best available option because it relies on clinically verified data rather than patient-reported outcomes.

Interpretation

Validation studies of diagnostic algorithms in other primary care databases show similar results to our findings. For instance, de Burgos-Lunar et al[21] conducted a validation study of diabetes mellitus and hypertension diagnoses in primary health care electronic records and found that both sensitivity and specificity were 99.5% for diabetes and were 85% and 97%, respectively, for hypertension. These results are consistent with our findings (100% sensitivity and 99% specificity for diabetes; 83% sensitivity and 98% specificity for hypertension). Another study also showed similar results for hypertension diagnosis, with sensitivity of 86% and specificity of 88%.[22]

It is difficult to compare results from validation studies that use different methodological approaches or that are based on administrative data. For example, one validation study used a 1:3 case-control design to assess the accuracy of COPD case-finding algorithms in administrative billing data.[23] The results from this study estimated a sensitivity of 61% and specificity of 82% (compared with our validation study, which indicated 41% sensitivity and 99% specificity). Another validation study for a diabetes diagnostic algorithm using administrative data yielded a sensitivity of 90% and specificity of 92%.[9] Finally, a large validation study of the Discharge Abstract Database evaluated the accuracy of administrative billing data. The accuracy of diagnostic coding was variable, yielding sensitivities of 68%, 57%, and 75% for COPD, diabetes, and hypertension, respectively.[8]

The variability in the accuracy of diagnostic coding for different conditions was also consistent with our findings. There are a number of reasons why the diagnostic algorithms might vary by condition. One such reason seems to be that diabetes and hypertension diagnoses are based on readily available and objective data, such as fasting glucose levels and blood pressure readings. Objective data for diagnosing the other 3 conditions are not as readily available. For example, spirometry for COPD diagnosis was underutilized in clinical practice (only 32% of patients newly diagnosed with COPD had undergone spirometry testing). More concerning are data that suggest that spirometric testing declines with increasing age.[24] Since our sample was stratified by age (90% of the selected patients were 60 years or older), we can assume that spirometric testing was underutilized in this age group. Thus, there is a lack of objective data to definitively ascertain diagnosis of COPD, which can lead to under-reporting. Another factor that can explain variability in diagnostic coding accuracy is the level of disease severity. For example, osteoarthritis can be present in patients who manage with minimal medical intervention (eg, use of over-the-counter medications) as well as in patients with debilitating pain that requires extensive management (eg, hip and knee replacement). In a preliminary analysis of discordant observations, chart abstractors reported that many true cases of osteoarthritis were found only because of a radiograph report that indicated osteoarthritic joints. Data from such reports are not readily available and currently are lacking from the CPCSSN diagnostic algorithms. This suggests that the osteoarthritis algorithm was able to capture more severe cases of osteoarthritis but missed the patients who had osteoarthritic joints yet managed with little medical intervention.

Implications

When considering the validity measures of diagnostic algorithms, it is important to appreciate the implications of tests with high sensitivity versus high specificity. High sensitivity indicates that the majority of true cases are identified, thus yielding few false negatives. Therefore, diagnostic tests with high sensitivity are useful if the purpose is to identify all or most of the true cases. In contrast, high specificity indicates that the majority of true non-cases are identified as such, thus yielding few false positives. Highly specific diagnostics are useful when the purpose is to identify only those who are true cases.

The findings show impressive specificity for all algorithms. High specificity suggests that patients who are identified as positive according to the algorithm are almost always true cases. Therefore, researchers who are interested in identifying a highly specific cohort of patients can use the current algorithms with the certainty that these algorithms will provide them with patients who are true cases.[25]

These results also show that the sensitivities of the algorithms vary by condition. The algorithm provided a perfect sensitivity for diabetes (100%) and a good sensitivity for hypertension (83%). High sensitivity for these 2 algorithms indicates that the majority of true cases are identified. In the diabetes example, all true cases were identified by the algorithms. High sensitivity is useful for policymakers who are interested in finding the prevalence of conditions within a certain population. This information can then be used to plan and allocate resources that are necessary in managing chronic conditions. However, the sensitivities for the other 3 algorithms were low (45%, 41%, and 39% for osteoarthritis, COPD, and depression, respectively).

It is important to underscore that sensitivity and specificity function in conjunction. For example, a highly specific yet poorly sensitive algorithm would eliminate false positives, but it may not represent the overall sample of cases.[25] Similarly, a highly sensitive yet poorly specific algorithm may capture all true cases but would also falsely include many noncases. Therefore, it is important to develop an algorithm that maximizes both sensitivity and specificity.

Conclusions

The diagnostic algorithm for diabetes demonstrates near-perfect accuracy, with 100% sensitivity and 99% specificity. Similarly, the algorithm for hypertension diagnosis demonstrates adequate accuracy, with 83% sensitivity and 98% specificity. Thus, these algorithms can be used for research and policymaking purposes. The diagnostic algorithms for the other 3 conditions demonstrate near-perfect specificity (100%, 99%, and 97% for osteoarthritis, COPD, and depression, respectively). However, future studies are needed to explore ways to enhance sensitivities for these 3 algorithms.

We thank research associates Suzanne Biro and Ashlynn Dundon for their contributions to data collection. We also thank data managers Ken Martin, Lorne Kinsella, and Sarah Sabri for their technical support for this project.

References

1. Schoen C, Osborn R, Doty MM, Bishop M, Peugh J, Murukutla N. Toward higher-performance health systems: adults' health care experiences in seven countries, 2007. Health Aff (Millwood) 2007;26:w717–34.
2. Dai S, Robitaille C, Bancej C, Loukine L, Waters C, Baclic O. Executive summary–report from the Canadian Chronic Disease Surveillance System: hypertension in Canada, 2010. Chronic Dis Can 2010;31:46–7.
3. Public Health Agency of Canada. Report from the National Diabetes Surveillance System–Diabetes in Canada, 2009. Ottawa: Public Health Agency of Canada; 2009.
4. Public Health Agency of Canada. The human face of mental health and mental illness in Canada, 2006. Ottawa: Public Health Agency of Canada; 2006.
5. Public Health Agency of Canada. Life and breath: respiratory disease in Canada. Ottawa: Public Health Agency of Canada; 2007.
6. Lagacé C, Public Health Agency of Canada. Life with arthritis in Canada: a personal and public health challenge. Ottawa: Public Health Agency of Canada; 2010.
7. Birtwhistle R, Keshavjee K, Lambert-Lanning A, et al. Building a pan-Canadian Primary Care Sentinel Surveillance Network: Initial development and moving forward. J Am Board Fam Med 2009;22:412–22.
8. Juurlink D, Institute for Clinical Evaluative Services. Canadian Institute for Health Information Discharge Abstract Database: a validation study [report online]. June 2006. Available from: http://www.ices.on.ca/file/cihi_dad_reabstractors_study.pdf. Accessed January 5, 2013.
9. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002;25:512–6.
10. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: use of electronic medical records for health outcomes research: a literature review. Med Care Re Rev 2009;66:611–38.
11. Kerby J, Keshavjee K, Holbrook A. Comparison of diagnostic codes in a clinical-research database and an administrative database. Proc AMIA Symp 2000;1045.
12. Broemeling A-M, Watson DE, Prebtani F. Population patterns of chronic health conditions, co-morbidity and healthcare use in Canada: implications for policy and practice. Healthc Q 2008;11:70–6.
13. Wurst KE, Ephross SA, Loehr J, Clark DW, Guess HA. The utility of the general practice research database to examine selected congenital heart defects: a validation study. Pharmacoepidemiol Drug Saf 2007;16:867–77.
14. Thomas SL, Edwards CJ, Smeeth L, Cooper C, Hall AJ. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general

practice research database? Arthritis Rheum 2008; 59:1314–21.
15. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. Fam Pract 2004;21: 396–412.
16. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;69:4–14.
17. Goulet JL, Erdos J, Kancir S, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Med Care 2007;45:73–9.
18. Hurdle JF, Weir CR, Roth B, Hoffman J, Nebeker JR. Critical gaps in the world's largest electronic medical record: ad hoc nursing narratives and invisible adverse drug events. AMIA Annu Symp Proc 2003;309–12.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.
20. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 2007;115:654–7.
21. de Burgos-Lunar C, Salinero-Fort MA, Cardenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. BMC Med Res Methodol 2011;11:146.
22. Duggan S, Aylett MJ, Eccles M, Ford GA. Defining hypertension in older people from primary care case notes review. J Hum Hypertens 1997;11:193–9.
23. Mapel DW, Frost FJ, Hurley JS, et al. An algorithm for the identification of undiagnosed COPD cases using administrative claims data. J Manag Care Pharm 2006;12:457–65.
24. Han MK, Kim MG, Mardon R, et al. Spirometry utilization for COPD: how do we measure up? Chest 2007;32:403–9.
25. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. J Clin Epidemiol 2012;65:343–49.e2.