A VALIDATION STUDY OF COMPUTER-BASED DIAGNOSTIC ALGORITHMS FOR CHRONIC DISEASE SURVEILLANCE

by

AMJED KADHIM-SALEH

A thesis submitted to the Department of Community Health & Epidemiology in conformity with the requirements for the degree of Master of Science

> Queen's University Kingston, Ontario, Canada July, 2012

Copyright © Amjed Kadhim-Saleh, 2012

Abstract

Background: Chronic conditions comprise a significant amount of healthcare utilization. For example, people with chronic diseases account for 51% of family physician encounters. Therefore, diagnostic algorithms based on comprehensive clinical records could be a rich resource for clinicians, researchers and policy-makers. However, limitations such as misclassification warrant the need for examining the accuracy of these algorithms.

Purpose: To investigate and enhance the accuracy of the diagnostic algorithms for five chronic diseases in the Canadian Primary Care Sentinel Surveillance Network.

Methods: DESIGN: A validation study using primary chart abstraction. SETTING: A stratified random sample of 350 patient charts from Kingston practice-based research network. OUTCOME MEASURES: Sensitivity and specificity for the diagnostic algorithms. ANALYSIS: A multiple logistic regression model along with the receiver operating characteristic curve was employed to identify the algorithm that maximized accuracy measures.

Results: The sensitivities for diagnostic algorithms were 100% (diabetes), 83% (hypertension), 45% (Osteoarthritis), 41% (COPD), and 39% (Depression). The lowest specificity was 97% for depression. A data-driven logistic model and receiver-operating characteristic curve improved sensitivity for identifying hypertension patients from 83% to 88% and for osteoarthritis patients from 45% to 81% with areas under the curve of 92.8% and 89.8% for hypertension and osteoarthritis, respectively.

ii

Conclusion: The diagnostic algorithms for diabetes and hypertension demonstrate adequate accuracy, thus allowing their use for research and policy-making purposes. A multivariate logistic model for predicting osteoarthritis diagnosis enhanced sensitivity while maintaining high specificity. This approach can be used towards further refining the diagnostic algorithms for other chronic conditions.

Co-Authorship Statement

This thesis presents research conducted by Amjed Kadhim-Saleh, under the supervision of Dr. Michael Green and Dr. Duncan Hunter. The study protocol was developed by Amjed Kadhim-Saleh with feedback from Michael Green, Tyler Williamson and Duncan Hunter. Michael Green, Richard Birtwhistle and Neil Drummond have contributed to the original conception of the study design. Data collection tool was developed by Amjed Kadhim-Saleh with feedback from Michael Green, Richard Birtwhistle, and Tyler Williamson. Data collection was conducted by Amjed Kadhim Saleh and by research associates at the Centre for Studies in Primary Care, specifically Suzanne Biro and Ashlynn Dundon. All data analyses were conducted by Amjed Kadhim-Saleh with feedback from Tyler Williamson and Michael Green. Amjed Kadhim-Saleh wrote the manuscripts and received feedback from Duncan Hunter, Michael Green and Tyler Williamson. Richard Birtwhistle also contributed content advice.

Acknowledgements

I am extremely grateful for everyone who has contributed to this thesis. First, I would like to express my gratitude for my supervisors, Dr. Michael Green and Dr. Duncan Hunter, who have provided me with tremendous support and guidance. Thank you so much for investing many hours in my training and for being wonderful mentors. Secondly, I am grateful to work closely with Dr. Tyler Williamson who has been a tremendous help. Tyler has been my mentor throughout every step of the thesis process and has dedicated countless hours towards this project. Third, I would like to acknowledge the support of Dr. Richard Birtwhistle who has also contributed to this thesis at various stages.

I am also grateful to the staff at the Centre for Studies in Primary Care who have contributed to data collection, especially Ashlynn Dundon and Suzanne Biro. I am also thankful to Ken Martin, Lorne Kinsella, Sarah Sabri and Shahriar Khan. I would also like to acknowledge the encouragement and support from Faculty and Staff at the Department of Community Health and Epidemiology. Special thanks to Erin Burnett, Lee Watkins and Gerri Lyman. I am very thankful for Dr. Kristan Aronson and Dr. Miu Lam for providing me the opportunity to be a Teaching Assistant for Epidemiology and Biostatistics.

This thesis could not have been accomplished without the financial support of the Ontario Graduate Scholarship, R.S. McLaughlin Fellowship, and internal funding from the Centre for Studies in Primary Care. The project was also supported by funding from the Public Health Agency of Canada.

v

Thank you to Michelle Dimitris for being a wonderful friend. Special thanks to Raymond Fong and Colleen Savage for their support over the past two years. I am also thankful to all the other amazing friends at the Department of Community Health and Epidemiology.

I would also like to thank Yilin Chen for her continuous love and for helping me achieve my goals. My two brothers, Muhanad and Ahmed, have been a constant joy in my life. Finally, I am forever indebted to my parents, Saad Saleh and Sukaina Radhi, who have never stopped believing in me. There are no words to describe their unconditional love and not enough words to express my gratitude for them.

Table of Contents

Abstractii
Co-Authorship Statementiv
Acknowledgementsv
Table of Contentsvii
List of Tablesx
List of Figuresxi
Chapter 1: General Introduction1
General Overview1
Limitations of electronic medical records-database2
Purpose
Background & Rationale4
The use of primary care databases in the surveillance of chronic diseases
Primary Care Databases in Other Countries6
Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Database
CPCSSN Diagnostic Algorithms7
Summary of Literature: Validation Studies of Primary Care Databases
Rationale9
Objectives
Thesis Organization
References
Chapter 2: Manuscript 1 17
Abstract
Introduction19
Methods
Design
Study population23
Definition of variables and data collection24
CPCSSN Diagnostic Algorithms24
Primary Chart Abstraction

Analysis
Results
Discussion
Summary of Main Findings
Strengths and limitations
Interpretation
Implications
Conclusions
References
Tables
Figures 44
Chapter 3: Manuscript 2 45
Abstract
Introduction
Methods 49
Design
Study population
Definition of variables and data collection49
CPCSSN Diagnostic Algorithms 49
Primary Chart Abstraction
Analytical strategy
Results
Discussion
Summary of Main Findings55
Comparison with Existing Literature
Strengths and Limitations57
Implications
Conclusions
References

Tables	65
Figures	69
Chapter 4: General Discussion	72
Summary of Study	72
General Strengths and Limitations	73
Overall Challenges	74
Implications	75
Conclusions	77
References	78
Appendices	79
Appendix 1. Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Disease Definitions	79
Appendix 2: Chart Abstraction Form	81
Appendix 3: Research Ethics Board Approval Letter	101

List of Tables

Chapter 2

Table 1. Sample demographics	41
Table 2. Patients' classification summary according to the gold standard	42
Table 2. Validation results of the five chronic conditions	43

Chapter 3

Table 1. Validation results of the Osteoarthritis diagnostic algorithm	.65
Table 2. Summary of the Univariate Analysis	.66
Table 3. Summary of the Multivariate Logistic Regression	67
Table 4. Results of the final model	67
Table 5. Testing the osteoarthritis diagnostic algorithm	68

List of Figures

Chapter 2

Figure 1 A visual re	presentation of the stud	ly design	11
Figure 1. A visual re	presentation of the stud	y uesign	44

Chapter 3

Figure 1. Receiver Operating Characteristic curve of the final model	69
Figure 1. Receiver Operating Characteristic curve of the re-created and tested model	70
Figure 1. Receiver Operating Characteristic curve of the sensitivity analysis	71

Chapter 1: General Introduction

General Overview

Chronic diseases are defined by the World Health Organization as diseases of long duration and generally slow progression. They are the leading cause of mortality in the world, representing 63% of all deaths (1). Recent estimates suggest that 46% of adult Canadians suffer from one or more of seven common chronic diseases (2).

In this study, five chronic conditions were chosen to be validated in a primary care electronic medical records-database: type 2 diabetes, hypertension, osteoarthritis, chronic obstructive pulmonary disease (COPD), and depression. These conditions comprise a significant portion of the morbidity and mortality burden, affecting 6 million Canadians with hypertension (3), 2 millions with diabetes (4), 1.2 million with major depression (5), more than 750,000 adults with COPD (6), and more than 3 million with osteoarthritis (7). It is important to note that the five index conditions often co-exist, such as hypertension and diabetes (3). According to data from the Canadian Chronic Disease Surveillance System, 5.1% of Canadians (about 1 million) aged 20 years and older were living with both diagnosed diabetes and hypertension in 2006-07 fiscal year (3).

In addition to contributing to morbidity burden, these conditions account for a significant amount of healthcare utilization, especially in the primary care setting. For instance, people with one or more chronic conditions account for 51% of family physician encounters (8). People with these chronic conditions are also more likely to have a regular doctor (92-96%) than those without these conditions (82%) (8). Furthermore, patients with a single index condition used 1.5 times the number of family physician encounters compared to those with no conditions (8).

The data above suggest that index conditions are managed at the primary care level, thus indicating that comprehensive clinical records collected by primary care physicians through electronic medical records (EMR) could be a rich resource for examining the impact of management and preventive strategies. Clinical records at the primary care level could be used to identify persons with chronic diseases, allowing researchers and policy makers to track diagnosis, treatment and management of chronic conditions.

Limitations of electronic medical records-database

In order to use primary care databases, it is important to investigate the quality of the data. There are a number of factors that contribute to diagnostic inaccuracy and incompleteness, thus highlighting the need for a validation study to quantitatively assess and enhance the accuracy. These factors include: misclassification, missing data, lack of standardization, and data usability limitations.

Misclassification can occur through various mechanisms. First, a number of studies have shown that computer-based electronic patient records often cannot differentiate between complex conditions, such as rheumatoid arthritis and osteoarthritis (9). Secondly, conditions with more subjective criteria are often under-diagnosed, such as mental illnesses (10). Thirdly, computer-based medical records can be misleading when a tentative diagnosis is classified as a definite one (11), thus mislabeling a patient as having the index condition when in fact the disease is absent. Misclassification can also occur when the computer-based algorithm excludes less severe cases that only require minimal treatment (11).

Missing data presents a serious limitation to the accuracy and completeness of computerbased diagnostic algorithms because these algorithms often lack important information, such as test results, referrals, and specialists' reports (11). The utilization of external health care may also be missing from the computer records and thus adds another limitation to the practicebased electronic medical records (12).

Another limitation in the use of practice-based electronic medical records is the lack of standardization across the different database platforms (12, 13). Several studies have indicated that the level of accuracy and completeness varies across the different systems (12). Even within the same EMR system, the level of accuracy and completeness can vary across the different providers and sites (10, 14, 15). This finding underscores the need for conducting validation studies for each different network or EMR system. Finally, there are a number of limitations to data usability, such as the inability to code certain fields (e.g. physician notes) (12, 16).

The limitations discussed above (misclassification, missing data, lack of standardization and data usability) warrant the need for a validation study to examine how these limitations impact the validity of using diagnostic-algorithms in primary care electronic medical records-databases.

Purpose

To validate the diagnostic algorithms for five chronic diseases in the Canadian Primary Care Sentinel Surveillance Network; to determine the impact of changes in diagnostic algorithms on accuracy measures.

Background & Rationale

This section will describe the use of primary care databases in the surveillance of chronic diseases with emphasis on the Canadian primary care context.

The use of primary care databases in the surveillance of chronic diseases

The concept behind surveillance has been widely used for infectious diseases. More recently, sentinel surveillance has been applied to major chronic diseases. Using information from a small group that represents a larger group, the results can be applied to understand the patterns of disease prevalence or risk factors. This has been used in the Public Health Agency of Canada (17).

Currently available information about chronic diseases at the national level is based on databases such as hospital discharge information, mortality data, 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 deaths or hospitalizations; unreliability of selfreported surveys especially when patients are not aware that they have the condition or if they are unwilling to report it (18). A large validation study of the Discharge Abstract Database concluded that coding of comorbidities that are present prior to admission or conditions that develop during hospital stays was very poor (14). For example, the complete agreement between the original record and re-abstracted record was only 3.7% for prior-to-admission COPD and 8% for type 2 diabetes (14). This study indicated that the sensitivity of some chronic conditions are relatively poor, such as a sensitivity of 68% for COPD (95% Confidence Interval: 64-72%), and 57% for type 2 diabetes (95% Confidence Interval: 50-64%) (14). Surveys are also limited in their use for ongoing surveillance due to financial burden (19).

At the provincial level, billing data on physician services and drug utilization provide a source of data, but it is very limited in the depth of information because administrative databases are created for financial management rather than research purposes (12). When compared against a clinical-research database, administrative data had only 20% agreement (20).

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 (21). Primary care databases can provide a wealth of information through examining various risk factors, medications, patient and provider characteristics, and diseases. Other benefits include the long-duration of follow-up, and the maximization of patient recruitment since patient consent is not required, given de-identified data and strict patient confidentiality and privacy measures (21). Also, the use of clinician-verified data minimizes

recall bias by the patient (21). In a retrospective study, researchers compared findings from randomized controlled trials and findings from the United Kingdom Clinical Practice Research Datalink (replicating previously performed trials) and found no significant differences between the two sources (22). This finding suggests that primary care databases can be even used to investigate therapeutic effectiveness. Moreover, an electronic medical database provides the opportunity to assimilate a large cohort, at relatively low cost, yielding an adequate power to detect small differences. Finally, randomized controlled trials often have strict protocols that may not be generalizable to the larger population and may not reflect the real-world application of therapy (12). In contrast, routine data collected through a primary care database has been shown to be cost-effective for conducting epidemiological research (23, 24).

Primary Care Databases in Other Countries

The use for primary care databases is not limited to Canada only. Other countries have established such databases, which have enabled researchers to investigate various conditions and policy makers to make more informed decisions. For example, the Clinical Practice Research Datalink in the United Kingdom (25) has been widely utilized. Over 890 peer reviewed publications were based on this database, covering a wide range of research areas from drug safety and effectiveness to genetic linkage and randomization studies, thus demonstrating its vast contribution to clinical research (25). Another example is the Netherlands Information Network of General Practice (26) which has also been widely used within the Netherlands as well as internationally. In Canada the National Research System (NaReS) of the College of Family Physicians of Canada has operated successfully for many years using an encounterreporting methodology (27).

Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Database

CPCSSN is an Electronic Medical Records-based information system for chronic disease surveillance (18) that was established in April 2008. Funded by the Public Health Agency of Canada (PHAC), it brings together sentinel practices in 10 practice-based research networks across the country (located in Alberta(2), Manitoba(1), Ontario(3), Quebec(1), British Columbia (1), and Atlantic Canada(2)) and institutional partners including academic research centres and departments, the College of Family Physicians of Canada and the Canadian Institute for Health Information. The central repository site is housed at the High Performance Virtual Computing Lab at Queen's University. Longitudinal data is extracted from the Electronic Medical Records of participating practices every three months. There are currently over 200,000 patients included in the database with 2-3 years of data extraction already available. Information contained in the database include: network and provider identifiers, de-identified patient demographics, encounter date and type, health conditions, risk factors, referrals, lab investigations, procedures, and medications. The coding system used for diagnosis is based on International Statistical Classification of Diseases and Related Health Problems (Version 9). In addition to providing important surveillance information to the Public Health Agency of Canada, CPCSSN data is available to qualified researchers for use in addressing important research questions in primary care including studies of the quality of care and impact of delivery models on the management of chronic diseases.

CPCSSN Diagnostic Algorithms

Case-finding diagnostic algorithms have been developed to identify patients with chronic conditions. The CPCSSN diagnostic algorithms have been published previously (18) and

are provided in **Appendix I**. These algorithms are based on various indicators, including: billing data, lab 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 lab tests (Hemoglobin A1C > 0.07, fasting blood sugar>7%). The International Statistical Classification of Diseases and Related Health Problems (version 9) was used along with certain drugs and/or positive test results to ascertain diagnosis. This system is the current coding standard for various EMRs in Canada; however, it may lack specificity for primary care (18).

Summary of Literature: Validation Studies of Primary Care Databases

There have been numerous validation studies conducted to evaluate the diagnostic accuracy of computer-based algorithms. Khan et al 2010 conducted a systematic review of all validation studies done on the United Kingdom-based General Practice Research Database (UK-GPRD) to assess the accuracy and completeness of diagnostic coding (24). A total of 49 papers were reviewed and results indicated that chronic diseases were generally well recorded when compared against the gold standard (General Practitioner questionnaire, hospital correspondence, or primary care medical records). Two main conclusions from the systematic review were that researchers should consider how well the index disease is recorded prior to planning a study and that they should consider how to optimize the identification of clinical events. This thesis targets both of these recommendations, with the first objective focusing on how well index conditions are identified, and the second objective on how to improve the accuracy of case-finding algorithms.

Herrett et al 2010 also conducted a systematic review of validation studies in the UK-GPRD (11). 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 in the validation studies. 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 (11). Also, there was insufficient detail in the sampling strategy, the percentage of missing data, and whether reviewers were blinded to the diagnosis (11). Secondly, the majority of validation studies were conducted on a highly selected cohort of patients; sampling only cases rather than selecting a random sample of both cases and non-cases. This stratified sampling lead to the inability to examine sensitivity and specificity measures, and thus only the positive predictive value (PPV) was calculated. Although informative, the PPV estimate is dependent on the prevalence of the condition, unlike sensitivity and specificity measures (11). Finally, the response rate for many of the validation studies was generally low ranging 55-100% (11).

Rationale

The summary of literature outlines a number of limitations in previous studies, including: insufficient quality of reporting, employment of test-based sample rather than random sampling, low response rate, and relatively small sample size. This thesis addresses these limitations by specifying the details of data to be reported (**see Appendix 2 for the chart abstraction tool**), using a random sample that validates both cases and non-cases, having full participation rate (all providers have already signed up to provide access to their patients' medical records) and having a relatively large sample size.

In addition to the limitations described above, there has been no validation study conducted on the CPCSSN database. As described earlier, there are four different layers that can lead to diagnostic inaccuracy and incompleteness: misclassification, missing data, lack of standardization due to the use of different EMR systems and having multiple providers, and data usability issues. As a result, it is necessary to conduct a validation study to understand the impact of these layers on the measures of accuracy.

Khan et al 2010 have correctly pointed out that while primary care databases are important for research, they are established for clinical rather than research purposes (24). Therefore, it is important to conduct studies to assess the quality of the data for the specific diseases of interest. While the comprehensive nature of electronic medical records provides multiple ways to identify patients, it is important to assess which method provides the best way to identify cases, with the highest possible sensitivity and specificity measures (12). Therefore, in the second objective, the impact of changes in diagnostic algorithms was assessed to explore which method provides better validity measures.

CPCSSN has already implemented many internal data quality checks and is working to ensure that the data extracted conforms to Canadian Institute for Health Information data quality framework. One aspect of this framework that requires additional research is comparison of the data extracted to a gold standard (28). This thesis fills this gap. A recent review of validation studies of the UK-GPRD highlights the importance of ensuring that data

quality and validity in EMR derived databases is adequate for the uses for which is it proposed (24). Improving our understanding of the quality of the data in EMR-based information systems such as CPCSSN will allow us to better appreciate the areas in which the data can be considered to be reliable and useful for both research and decision making, to understand the limitations of data acquired in this manner and to improve the quality of such data in the future.

Objectives

- To determine the sensitivity and specificity of diagnostic algorithms used to identify five chronic conditions (type 2 diabetes, hypertension, osteoarthritis, chronic obstructive pulmonary disease, and depression), compared with primary chart abstraction in the Canadian Primary Care Sentinel Surveillance Network
- To assess the impact of changes in these diagnostic algorithms on the sensitivity and specificity of the identification of these conditions

Thesis Organization

This thesis conforms to the framework provided by the "General Forms of Theses" as outlined by the School of Graduate Studies at Queen's University (29). The second chapter of this thesis is the first manuscript, which assesses the validity of Electronic Medical Records-based diagnostic algorithms for five chronic conditions in the Canadian Primary Care Sentinel Surveillance Network database. This manuscript has been prepared for submission to the Journal of American Board of Family Medicine. Chapter three is the second manuscript, which addresses objective 2. This manuscript applies a methodological approach towards enhancing and testing the validity of a new algorithm to improve accuracy for Osteoarthritis case-finding. This manuscript has been prepared for submission to the Journal of American Epidemiology. Chapter four contains general conclusions, discussion and a summary of the study.

References

1. World Health Organization. Health Topics, Chronic Diseases. Retrieved May 19, 2012 from the World Wide Web: http://www.who.int/topics/chronic_diseases/en/.

2. Schoen C, Osborn R, Doty MM, Bishop M, Peugh J, Murukutla N. Toward higherperformance health systems: adults' health care experiences in seven countries, 2007. Health.Aff.(Millwood). 2007 Dec;26(6):w717-34.

3. 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 Dec;31(1):46-47.

4. Public Health Agency of Canada. Report from the National Diabetes Surveillance System Diabetes in Canada, 2009. Ottawa :: Public Health Agency of Canada,; 2009.

5. Public Health Agency of Canada. The human face of mental health and mental illness in Canada, 2006. [Ottawa]: Public Health Agency of Canada; 2006.

6. Public Health Agency of Canada. Life and Breath: Respiratory Disease in Canada. 2007;

7. 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.

8. 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(3):70-76.

9. 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 Sep 15;59(9):1314-1321.

10. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. Family Practice. 2004 Aug;21(4):396-412.

 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 Jan;69(1):4-14.

12. 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. Medical Care Research and Review. 2009 Mar;66(6):611-638.

13. Thiru K, Hassey A, Sulliv F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003 May;326(7398):1070-0.

14. Juurlink D. Canadian Institute for Health Information Discharge Abstract Database: A Validation Study. 2006;

15. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Med Care. 2007 Jan;45(1):73-79.

16. 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;:309-312.

17. Public Health Agency of Canada. Chronic diseases. Available at: http://www.phacaspc.gc.ca/cd-mc/index-eng.php. Accessed July 5, 2011.

Birtwhistle R, Keshavjee K, Lambert-Lanning A, Godwin M, Greiver M, Manca D, et al.
 Building a Pan-Canadian Primary Care Sentinel Surveillance Network: Initial Development and
 Moving Forward. The Journal of the American Board of Family Medicine. 2009 Jul;22(4):412 422.

19. 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 Mar;25(3):512-516.

20. Kerby J, Keshavjee K, Holbrook A. Comparison of Diagnostic Codes in a Clinical-research Database and an Administrative Database. Proceedings of the AMIA Symposium. 2000;:1045.

21. 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 Aug;16(8):867-877.

22. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ. 2009 Jan 27;338:b81.

23. Jordan K, Porcheret M, Kadam UT, Croft P. The use of general practice consultation databases in rheumatology research. Rheumatology (Oxford). 2006 Feb;45(2):126-128.

24. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br.J.Gen.Pract. 2010 Mar;60(572):e128-36.

25. The Clinical Practice Research Database [Internet]. Cited 2012, June 7; Available from: http://www.cprd.com/intro.asp

26. Netherlands Institute for Health Services Research. Databases and information systems:
Netherlands Information Network of General Practice [Internet]. 2011 Jul 19 [cited 2011 Jul 19];Available from: http://www.nivel.nl/linh/

27. The College of Family Physicians of Canada. National Research System (NaReS) [Internet]. 2011 Jul 19 [cited 2011 Jul 19]; Available from: <u>http://www.cfpc.ca/nares/</u>

28. Canadian Institute for Health Information, The CIHI Data Quality Framework, 2009 (Ottawa, Ont.: CIHI, 2009).

29. School of Graduate Studies and Research. (2011). General Forms of Theses. Kingston,ON: Queen's University.

Chapter 2: Manuscript 1

Title: Validation of the diagnostic algorithms for five chronic conditions in the Canadian Primary

Care Sentinel Surveillance Network

Authors:

Kadhim-Saleh A., BHSc MSc (Cand); Green M., MD, MPH, CCFP; Williamson T., PhD;

Hunter D., PhD; Birtwhistle R., MD MSc FCFP

Affiliations:

Mr. Kadhim-Saleh and Dr. Hunter are affiliated with the Department of Community Health and Epidemiology, Queen's University. Drs. Green, Williamson and Birtwhistle are affiliated with the Department of Community Health and Epidemiology and the Department of Family Medicine, Queen's University.

Correspondence: Richard Birtwhistle, MD MSc CCFP FCFP

Department of Family Medicine, Queen's University

220 Bagot St, Kingston, ON K7L 5E9; telephone 613 533-9300, extension 73934;

Email: birtwhis@queensu.ca

Keywords: practice-based research network; health informatics; chronic diseases

This manuscript has been formatted for submission to the peer-reviewed Journal of the

American Board of Family Medicine

Manuscript word count (text only, excluding abstract and references): 4,235 words Manuscript page count (includes title page, abstract, references, tables, and figures): 25 pages

Abstract

OBJECTIVE: To assess the validity of electronic medical records-based diagnostic algorithms for five chronic conditions.

DESIGN: 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 (COPD), and Depression. Information was collected on billing, lab tests, notes, specialist and hospital reports, and physiological data. PARTICIPANTS: An age-stratified random sample of 350 patient charts was selected from Kingston, Ontario. Approximately 90% of those charts were allocated to people 60 years or older.

RESULTS: Three hundred and thirteen patient charts were included in the study. The mean age of patients was 68 years old and 52% were female. The inter-rater reliability had 92% complete agreement and 89.3% kappa statistic. The sensitivities of algorithms were 100% (diabetes), 83% (hypertension), 45% (osteoarthritis), 41% (COPD), 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 three conditions require further refinement to attain better sensitivities.

Introduction

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 seven common chronic diseases (1). These conditions affect 6 million Canadians with hypertension (2), 2 millions with diabetes (3), 1.2 million with major depression (4), more than 750,000 adults with Chronic Obstructive Pulmonary Disease (COPD) (5), and more than 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 comorbidities was poor (8). For example, the complete agreement between the original record and reabstracted record was only 3.7% for prior-to-admission COPD and 8% for diabetes (8). Surveys are also limited in their use for ongoing surveillance due to added financial burden (9).

At the provincial level, billing data on physician services may provide a source of data, but it is limited in the depth of information because administrative data is 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 policy-makers. 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 across the country and institutional partners including academic research centres and departments, the College of Family Physicians of Canada and the Canadian Institute for Health Information (7). Longitudinal data is extracted from the participating practices every three months and include the following information: network and provider identifiers, de-identified patient demographics, encounter date and type, health conditions, risk factors, referrals, lab investigations, procedures, and medications. There are currently over 300,000 patient records included in the database with 2-3 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, lab test results, and medications to ascertain diagnoses in the database. For example, patients with diabetes can be identified from existence of billing data (code 250.X), medications (insulin, glyburide, metformin) and lab tests (Hemoglobin A1C >0.07, fasting blood sugar>7%).

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

Misclassification can occur due to difficulty in differentiating between complex conditions, under-reporting 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 utilization 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 (e.g. physician notes) (10, 18).

Many validation studies have been conducted to evaluate the accuracy of computerbased diagnostic algorithms. Herrett et al 2010 conducted a systematic review of validation studies in the United Kingdom-Clinical Practice Research Datalink (UK-CPRD) (16). 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. Also, there was insufficient detail in the sampling strategy, the percentage of missing data, and whether reviewers were blinded to diagnosis. Secondly, the majority of validation studies were conducted on a highly selected cohort of patients that only considered cases rather than selecting a random sample of both cases and non-cases. This case-stratified sampling lead 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 55-100%.

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, utilizing a random sample that will validate both cases and non-cases, and having full participation rate. Improving the understanding of the quality of the 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 is to

assess the validity of EMR-based diagnostic algorithms for five 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 visual presentation of comparing the level of concordance between two data sources, CPCSSN database and the Primary Chart Abstraction of patients' records. This was a pilot study for the larger CPCSSN validation project, which will take place at each of the 10 practice-based research networks. 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% or less, an acceptable level of precision.

[Insert Figure 1 here]

Study population

Patients who attended the Kingston (Ontario) practice-based research network were included in the study. An age-stratified random sample of 350 patient charts was selected, with 90% of these charts allocated to people 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. The CPCSSN diagnostic algorithms have been published previously (7). These algorithms are based on various indicators, including: billing data, lab 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 lab tests (Hemoglobin A1C > 0.07, fasting blood sugar>7%). The International Statistical Classification of Diseases and Related Health Problems (version 9) was used along with certain drugs and/or positive test results to ascertain diagnosis (7). This system is the current coding standard for various EMRs in Canada; however, it may lack specificity for primary care (7). CPCSSN extracts data from sentinel practices every three months. For the validation study, data from the third quarter of 2011 was used to ascertain the diagnoses of patients who were included. This step was conducted after the completion of the primary chart abstraction.

Primary Chart Abstraction

The gold standard in this study was the primary chart audit of electronic records undertaken by three trained and experienced researchers. Charts were reviewed to determine if patients 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 practice-based research network. 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, family physician notes, and specialist and hospital reports were collected. The chart abstraction was done electronically using Microsoft Access Database. Chart abstractors were blinded to the CPCSSN diagnoses. The first 10 charts reviewed by research associates were re-abstracted independently by a second associate to ensure adequate data quality and consistency. Reviewers then consulted with a clinician (MG, RB) who arbitrated on cases of disagreement.

Patients were classified into one of three categories: 1) confirmed positive; 2) confirmed negative; 3) presumed negative. 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 who had none of the indicators for diabetes and a normal level of hemoglobin A1C was classified as a true-negative. Patients were classified as "presumed negative" when there were neither indicators for the disease nor any tests done to rule out the condition. Patients were also flagged 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 Statistical Analysis Software 9.2 Version (26). Sensitivity, specificity, positive predictive and negative predictive values 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 diagnosticalgorithms versus primary chart abstraction. The level of inter-rater 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 chart 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 chart abstraction) by the total number of true noncases. When calculating sensitivity and specificity, patients who were classified as "confirmed negative" or "presumed negative" were considered to lack the index disease according to chart abstraction. A proportion test was conducted to determine whether the sampling strategy of selecting different age groups affected the sensitivity and specificity. In order 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. A descriptive analysis was conducted to determine the characteristics of patients who were identified as true cases according to chart abstraction, but were classified as false negative by the CPCSSN algorithms. The study protocol was approved by the Queen's Health Sciences Research Ethics Board.

Results

The study population was 313 patients. Thirty seven patient charts were excluded from the analysis due to lack of data (non-rostered 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 old and 52% were female. In a sample of 10 charts that were reviewed independently, the exact percent agreement between reviewers was: 100% (diabetes, osteoarthritis), 90% (COPD, depression) and 80% (hypertension). Thus, the overall complete agreement between chart abstractors was 92%. The overall Kappa Statistic was 89.3%. In approximately 5% of patient charts, a clinician (MG) was consulted on the appropriate diagnoses. The clinician examined the full chart in order to reach a decision.

Based on chart abstraction, approximately 80% of patients had at least one of the five 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, including the prevalence of each condition, as well as the prevalence of multiple morbidities.

[Insert Table 1 here]

Table 2 summarizes the classification of patients' diagnoses according to the gold standard into three categories: confirmed positive, presumed negative, and confirmed negative. Table 3 provides details on the concordance and discordance between the two data sources, chart abstraction and CPCSSN algorithms. It also provides details on the sensitivity,

specificity, positive predictive value, and negative predictive value, along with the exact 95% confidence intervals. The proportion test revealed no significant age effect on sensitivity and specificity estimates at the 0.05 level. 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 the chart abstraction, there were 16.9% of patients with type 2 diabetes. The diagnostic algorithm for identifying patients with diabetes had 100% sensitivity, 99% specificity, 95% PPV, and 100% NPV.

There were approximately 57.8% of patients with hypertension. The diagnostic algorithm for identifying patients with hypertension had 83% sensitivity, 98% specificity, 98% PPV, and 81% NPV.

There were approximately 46% of patients with osteoarthritis. The diagnostic algorithm for identifying patients with osteoarthritis had 45% sensitivity, 100% specificity, 100% PPV, and 68% NPV.

There were approximately 9.2% of patients with COPD. The diagnostic algorithm for identifying patients with COPD had 41% sensitivity, 99% specificity, 80% PPV, and 94% NPV.

Finally, there were approximately 21% of patients with depression. The diagnostic algorithm for identifying patients with depression had 39% sensitivity, 97% specificity, 79% PPV, and 86% NPV.

[Insert Table 2 here]

[Insert Table 3 here]

The descriptive analysis of true cases that were missed by the CPCSSN algorithms revealed that the vast majority of these missing cases were identified through physician notes, such as encounter diagnosis. Out of the 31 true hypertension cases that were missed by CPCSSN, 28 were identified by the chart abstractors through physician notes. Similarly, 87% of missing Osteoarthritis cases, 76% of missing COPD cases, and 98% of missing depression cases were identified through physician notes. For Osteoarthritis, the second and third characteristics that were mostly likely used to identify missing cases were clinical images (39%) and referrals (30%). For Depression, the second most likely characteristic that was used to identify missing cases was referrals (25%).

Discussion

Summary of Main Findings

The CPCSSN project was initiated to provide ongoing surveillance of chronic conditions. Additionally, the CPCSSN data could be used by researchers to identify a cohort of patients and examine the effectiveness of prevention and management strategies. Policy-makers could also use this data to plan and allocate resources needed to manage chronic conditions. The findings suggest that the specificity for all five conditions was very high (lowest of 97% for depression). Thus, the diagnostic algorithms are highly specific and yield very few false positive cases. Sensitivities of the CPCSSN algorithms, in contrast, varied considerably among the five conditions. The sensitivity of diagnostic algorithms for diabetes (100%) and hypertension (83%) was adequate. Thus, the majority of true diabetes and hypertension cases 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 under-estimating the true prevalence of these three conditions.

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 rate. For instance, a validation study using an administrative database in Ontario had a response rate of only 11% (9). A low response rate may have lead to selection bias and a lack of generalizability because responding practices are systematically different from non-responding practices (16). In this study, medical records from all randomly selected practices were examined to assess presence or absence of index conditions.

The sample size was relatively large compared to similar studies done in the United Kingdom Clinical Practice Research Datalink (14). Also, the random sampling technique utilized 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 positive predictive value 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 non-cases, this study ensured that all patients (cases and non-cases) 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).

The limitations of this study ought to be considered before any conclusions can be drawn. One challenge was the use of different EMRs by research networks. All EMRs have different coding structures and 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 validity measures. The study could not assess the generalizability of findings across the CPCSSN database, since it examined accuracy at the Kingston network alone. In other words, data from one network may not represent the overall data quality. This study was a pilot project for the larger CPCSSN validation. There are at least two other similar projects being currently conducted at two sites in Alberta and Manitoba. Once these findings are known, validity measures can be compared across the different sites.

The chart abstraction was based on the electronic records available through the Kingston EMR platform entitled OSCAR. It is important to consider the level of completeness of these records to understand the limitations associated with this process. OSCAR records including physician notes and medications are available for all patients starting from 2004. Other data including test results, specialist and hospital correspondence were available from June 2010 for all patients. Prior to 2010, documents were imported into OSCAR selectively for some patients based on physicians' requests.

A number of data quality issues were identified during a feasibility project of CPCSSN along with remediation strategy. These issues are summarized previously (7) and include: dirty data (e.g. a misspelling), missing data, and inconsistent data. Some strategies included cleaning algorithms, training physicians to enter appropriate data, and using dates to determine latest status of risk factors.

The study was also limited to a retrospective analysis that relied on clinical records that have been already documented by different health providers. Thus, the documentation was done inconsistently and a longitudinal assessment could be considered more accurate than a retrospective design. It is also important to acknowledge the limitation that the study is subject to measurement error since a true gold standard is not available. Primary chart abstraction is an imperfect gold standard and thus is subject to measurement error (20).

Interpretation

Validation studies of diagnostic algorithms in other primary care databases show similar results to our findings. For instance, Carmen de Burgos-Lunar et al (2011) 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 85% sensitivity and 97% specificity for hypertension (21). These results are consistent with our findings (100% sensitivity and 99% specificity for diabetes; 83% sensitivity and 98% specificity for hypertension diagnoses). 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 to 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 Discharge Abstract Database evaluated the accuracy of administrative billing data. The accuracy of diagnostic coding was variable, yielding the following sensitivities: 68%, 57%, and 75% for COPD, diabetes, and hypertension, respectively (8).

The variability in diagnostic coding accuracy was consistent with our findings, where two algorithms were adequate in case-finding (high sensitivities for diabetes and hypertension diagnostic algorithms) while the other three algorithms (osteoarthritis, COPD, and depression) had near-perfect specificities with low sensitivities. There could be a number of reasons why the diagnostic algorithms varied according to 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 three conditions are not as readily available. For example, spirometry for COPD diagnosis was under-utilized in clinical practice (only 32% of newly diagnosed patients with COPD had undergone spirometry testing). More concerning is data that suggests that spirometric testing declines with increasing age (24). Since our sample was stratified by age where 90% of the selected patients were 60 years or older, we can assume that spirometric testing was under-

utilized in this age group. Thus, there is a lack of objective data to definitively ascertain diagnosis of COPD which can lead to under-reporting. The other reason that can explain the variability of diagnostic coding accuracy is the level of severity of disease. For example, osteoarthritis can present in a range of severity from patients who manage with minimal medical intervention (e.g. use of over-the-counter medications) to debilitating pain that requires extensive management (e.g. hip and knee replacement). In a preliminary analysis of discordant observations, chart abstractors reported that many true cases of osteoarthritis were only found because of an x-ray report that indicated osteoarthritic joints. Data from such reports are not readily available and are currently 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 high sensitivity versus high specificity values. High sensitivity indicates that the majority of true cases are identified, thus yielding very few false negatives. Therefore, high sensitivity diagnostics 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 very few false positives. Highly specific diagnostics are useful when the purpose is to *only* identify those who are true cases.

The findings show that the specificity for all algorithms is nearly perfect. 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 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 two 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 policy-makers 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 three algorithms were low (45%, 41%, 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 non-cases. Therefore, it is important to develop an algorithm that maximizes both sensitivity and specificity. This description seems to fit the algorithms for both diabetes and hypertension, thus allowing their use for research and policy-making purposes. The other three algorithms do have high specificity, but poor sensitivity. Thus, future studies are needed to explore ways to enhance the sensitivities of those three algorithms.

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 policy-making purposes. The diagnostic algorithms for the other three conditions demonstrate near-perfect specificity (100%, 99%, 97% for osteoarthritis, COPD, and depression, respectively). However, future studies are needed to explore ways to enhance sensitivities for these three algorithms.

References

1. Schoen C, Osborn R, Doty MM, Bishop M, Peugh J, Murukutla N. Toward higherperformance health systems: adults' health care experiences in seven countries, 2007. Health.Aff.(Millwood). 2007 Dec;26(6):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 Dec;31(1):46-47.

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; 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. The Journal of the American Board of Family Medicine. 2009 Jul;22(4):412-422.

8. Juurlink D. Canadian Institute for Health Information Discharge Abstract Database: A Validation Study. 2006;

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 Mar;25(3):512-516.

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. Medical Care Research and Review. 2009 Mar;66(6):611-638.

11. Kerby J, Keshavjee K, Holbrook A. Comparison of Diagnostic Codes in a Clinical-research Database and an Administrative Database. Proceedings of the AMIA Symposium. 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(3):70-76.

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 Aug;16(8):867-877.

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 Sep 15;59(9):1314-1321.

15. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. Family Practice. 2004 Aug;21(4):396-412.

 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 Jan;69(1):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 Jan;45(1):73-79.

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;:309-312.

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-174.

20. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007 Feb 6;115(5):654-657.

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 Oct 28;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 Mar;11(3):193-199.

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 Jul-Aug;12(6):457-465.

24. Han MK, Kim MG, Mardon R, et al. Spirometry utilization for COPD: how do we measure up? Chest 2007 Aug;132(2):403-409.

25. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. J Clin Epidemiol 2012 Mar;65(3):343-349.e2.

26. SAS Institute Inc., Version 9.2, Cary, NC.

Tables

Table 1. Sample demographics of patients included in the study from the Kingston Practice-Based Research Network

Demographics	Count (n=313)	Percentage
Male	150	48%
Female	163	52%
≥60 years old	282	90%
<60 years old	31	10%
Diabetes	53	17%
Hypertension	181	58%
Osteoarthritis	144	46%
COPD	29	9%
Depression	66	21%
At least one of the five chronic conditions	250	80%
Has two chronic conditions	94	30%
Has three chronic conditions	37	12%
Has four chronic conditions	17	5%

Condition	Confirmed Positive	Presumed Negative	Confirmed Negative		
Diabetes	29	272	12		
Hypertension	62	184	49		
Osteoarthritis	53	81	179		
COPD	181	70	62		
Depression	144	169	0		

Table 2. Patients' classification summary according to the gold standard (chart abstraction) for the five chronic conditions.

Condition	Chart +	Chart -	Chart +	Chart -	Sens. (95% CI)	Spec. (95% Cl)	PPV (95% CI)	NPV (95% CI)
	CPCSSN +	CPCSSN +	CPCSSN -	CPCSSN -				
Diabetes	53	3	0	257	100% (92-100)	99% (96-100)	95% (83-100)	100% (98-100)
Hyper- tension	150	3	31	129	83% (75-89)	98% (93-100)	98% (94-100)	81% (72-88)
Osteo- arthritis	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)

Table 3. Validation results of the case-finding diagnostic algorithm for the five chronic conditions.

*95% exact confidence intervals. PPV= Positive Predictive Value. NPV= Negative Predictive Value.

Figures

Figure 1. A visual representation of the study design. Computerized data collected through CPCSSN will be compared against the gold standard (primary chart abstraction). For each index condition, a CPCSSN protocol algorithm was used to determine whether or not a patient had the disease. In contrast, pre-specified criteria were used in chart abstraction to identify presence or absence of index conditions. The level of concordance will be compared between the two data sources.



Chapter 3: Manuscript 2

Title: Validating and enhancing an algorithm for identifying cases of osteoarthritis in a primary

care sentinel surveillance database

Authors:

Kadhim-Saleh A., BHSc MSc (Cand); Green M., MD, MPH, CCFP; Williamson T., PhD; Hunter D., PhD; Birtwhistle R., MD MSc FCFP

Affiliations:

Mr. Kadhim-Saleh and Dr. Hunter are affiliated with the Department of Community Health and

Epidemiology, Queen's University. Drs. Green, Williamson and Birtwhistle are affiliated with the

Department of Community Health and Epidemiology and the Department of Family Medicine,

Queen's University.

Correspondence: Richard Birtwhistle, MD MSc CCFP FCFP

Department of Family Medicine, Queen's University

220 Bagot St, Kingston, ON K7L 5E9; telephone 613 533-9300, extension 73934;

Email: birtwhis@queensu.ca

Keywords: Public health informatics; Osteoarthritis; Multivariate Analysis; ROC Curve

This manuscript has been formatted for submission to the peer-reviewed American Journal of

Epidemiology

Manuscript word count (excluding abstract and references): 3,626

Manuscript page count (including abstract, references, tables, and figures): 27 pages

Abstract

INTRODUCTION: Electronic Medical Records (EMRs) can be a rich source of data on management and prevention strategies for osteoarthritis. Algorithms that identify persons with osteoarthritis can be a useful resource for researchers and policy-makers.

PURPOSE: To investigate and enhance the accuracy of a diagnostic algorithm for chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN), using Osteoarthritis as an example.

METHODS: DESIGN: A retrospective analysis on the accuracy of EMR-based diagnostic algorithms. SETTING: A stratified random sample of 350 patient charts from Kingston, Ontario. OUTCOME MEASURES: Sensitivity and specificity for diagnostic algorithms. ANALYSIS: Using multivariate logistic regression and a Receiver Operating Characteristic (ROC) curve, a diagnostic algorithm was developed to maximize sensitivity and specificity.

RESULTS: The existing CPCSSN algorithm had 45% sensitivity, 100% specificity, 100% positive predictive value, and 68% negative predictive value. A data-driven model and ROC curve improved sensitivity for identifying osteoarthritis patients from 45% to 81% and maintained high specificity (90%).

CONCLUSION: The existing osteoarthritis algorithm had perfect specificity but poor sensitivity. A logistic model for predicting osteoarthritis demonstrated higher sensitivity; while maintaining high specificity and was validated in a sample dataset. This approach can be used towards further refining algorithms for other chronic conditions.

Introduction

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 seven common chronic diseases (1). Osteoarthritis is one of those conditions, affecting more than 3 million adult Canadians (2) and approximately 27 million adults in the United States (3).

Osteoarthritis is a condition characterized by cartilage deterioration and bone thickness that cause joint damage and stiffness. The possible risk factors for osteoarthritis include age, genetic predisposition, obesity, and prior joint injury (2). Early diagnosis is crucial for appropriate management strategies that reduce the impact of arthritis. These strategies range from self-management and education to counselling, rehabilitation therapy, weight reduction, medications, and surgery (2).

There is evidence to suggest that primary care physicians contribute significantly towards osteoarthritis patient care. For example, approximately 80% of people with arthritis have visited a primary care physician at least once in 2005-2006 (2). This data suggests that comprehensive clinical records collected by primary care physicians can be a rich resource for researchers interested in examining prevention and management strategies for osteoarthritis. Additionally, this resource can inform policy makers on issues of access and quality of care. The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) is one such source of clinical information on chronic diseases including osteoarthritis. It is an Electronic Medical Records (EMR)-based information system for chronic disease surveillance (4). It brings together sentinel practices in 10 practice-based research networks and institutional partners including academic

research centres, the College of Family Physicians of Canada and the Canadian Institute for Health Information. Longitudinal data is extracted from the participating practices every three months that include the following information: network and provider identifiers, de-identified patient demographics, encounter date and type, health conditions, risk factors, referrals, lab investigations, procedures, and medications. There are currently over 300,000 patients included in the database with 2-3 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, lab test results, and medications to ascertain diagnoses. For example, patients with osteoarthritis would be identified from the existence of billing data (code 715.X), risk factors (e.g. obesity and joint pain), referrals (e.g. orthopedic surgeon) and procedures (knee and hip replacement).

Previously, we conducted a validation study of the EMR-based diagnostic algorithms used by CPCSSN for identifying patients with diabetes, hypertension, osteoarthritis, chronic obstructive pulmonary disease (COPD), and depression (16). The results showed adequate accuracy for diabetes (100% sensitivity, 90% specificity) and hypertension (83% sensitivity, 98% specificity). The algorithms for the other three conditions had near-perfect specificities, but poor sensitivities (45%, 41%, and 39% for osteoarthritis, COPD, and depression, respectively). Poor sensitivity indicates that the diagnostic algorithms under-report the true prevalence of these conditions. In this paper, we report our efforts to enhance the sensitivity of these diagnostic algorithms by using osteoarthritis as an example. The objective of this study is to

identify characteristics that predict osteoarthritis diagnosis and develop a model that enhances the validity of osteoarthritis EMR-based diagnostic algorithm in the CPCSSN database.

Methods

Design

This study was a retrospective analysis on the accuracy of case-finding diagnostic algorithms in EMR-based information. Data collected through CPCSSN were compared against the gold standard (primary chart abstraction). The CPCSSN diagnostic algorithm was used to determine whether or not a patient has osteoarthritis. In contrast, pre-specified criteria were used in chart abstraction to identify presence or absence of osteoarthritis diagnosis. This was a pilot study for the larger CPCSSN validation project, which will take place at each of the 10 practice-based research networks. Age-stratified sampling was used to ensure that the disease prevalence would be sufficiently high. This, in turn, would ensure that the width of the 95% confidence interval would be maintained at 10% or less, an acceptable level of precision (5).

Study population

Patients who attended the Kingston practice-based research network were included in the study. An age-stratified random sample of 350 patient charts was selected, with 90% of these charts allocated to people 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. The CPCSSN diagnostic algorithms have been published previously (4). These algorithms are based on various indicators, including: billing, lab test results, and medications. CPCSSN extracts data from sentinel practices every three months. For our validation study and the retrospective analysis, we used data from the third quarter of 2011 to ascertain the diagnoses of patients. This step was conducted after completing the chart abstraction.

Primary Chart Abstraction

The gold standard was considered to be the primary abstraction of electronic records performed by three trained and experienced research associates. Charts were reviewed to determine if patients 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 use the EMR platform at the Kingston practice-based research network. 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, family physician notes, and specialist and hospital reports were collected. No limitations in time periods were set for collecting these data. However, most electronic records in our sample spanned approximately 7-8 years. The chart abstraction was done electronically using Microsoft Access Database. Chart abstractors were blinded to the CPCSSN diagnoses. The first 10 charts reviewed by research associates were reabstracted independently by a second associate to ensure adequate data quality and consistency. Reviewers then consulted with a clinician (MG, RB) who arbitrated on cases of disagreement.

Patients were classified into either confirmed positive for osteoarthritis or unsure. Confirmed positive classification was used when there was sufficient evidence to conclude that the patient was a true case. For example, a patient was classified as confirmed positive if the physician notes and imaging reports confirmed that the patient had osteoarthritis. Unsure classification was used when there were no indicators present for osteoarthritis. In cases of contradictory evidence, consultation with a clinician (MG) was sought to reach a decision.

Analytical strategy

All calculations were done using Statistical Analysis Software 9.2 Version (17). Two-sided *p* values were used. Sensitivity, specificity, positive predictive and negative predictive values were calculated along with exact 95% confidence intervals. A two-by-two contingency table was created specifying whether osteoarthritis was present or absent according to the CPCSSN diagnostic-algorithm versus primary chart abstraction. In addition, the level of inter-rater reliability was expressed as a Kappa statistic and percent agreement.

Sensitivity was calculated by dividing the number of records in which a diagnosis was present according to both sources (CPCSSN and chart 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 chart abstraction) by the total number of true noncases.

In order to enhance the accuracy measures of the osteoarthritis diagnostic algorithm, a method from a similar study that investigated algorithms accuracy in patients with rheumatoid arthritis and juvenile idiopathic arthritis in a large primary care database was employed (6). Firstly, univariate analyses were conducted to ascertain the association between each variable (e.g. procedures) and osteoarthritis diagnosis using Chi-Square Statistic and Odds Ratio (OR). Next, a multivariate logistic regression model was built by sequentially adding statistically significant variables (OR \geq 2 or p-value \leq 0.1) and keeping variables if they retained an association with osteoarthritis diagnosis (p-value \leq 0.2). The variable with the largest chi-square statistic was added first. Thirdly, a prediction score was derived by summing the log of the odds ratios for variables in the final model and each patient was assigned a predictive score based on whether he or she had osteoarthritis indicators. Then, a Receiver Operating Characteristic Curve was utilized to determine the best cut-off score that maximized sensitivity and specificity. Finally, the validity of this data-driven algorithm was tested in the whole sample first. Then, the full dataset was split randomly in half, re-creating the algorithm in one half and testing it in the other half (6).

Results

The study population included a total of 313 patients. Thirty seven patient charts were excluded from the analysis due to lack of data (non-rostered 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 old and 52% were female. Based on the chart abstraction, approximately 46% of patients had osteoarthritis (Table 1).

In a sample of 10 charts that were reviewed independently, the exact percent agreement between reviewers was 100% for osteoarthritis and the overall exact agreement was 92%. The overall Kappa Statistic was 89.3%. In approximately 6% of patient charts, a clinician (MG) was consulted on the appropriate diagnosis.

The existing CPCSSN algorithm used billing data and medical history to identify patients with osteoarthritis. Compared with the gold standard, this algorithm had 45% sensitivity, 100% specificity, 100% positive predictive value and 68% negative predictive value (Table 1).

[Insert Table 1 here]

Univariate Analysis: Ten potential predictors of osteoarthritis diagnosis were assessed, including billing (code 715.x), condition (i.e. osteoarthritis disease registry), analgesics, Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) (excluding naproxen and Acetylsalicylic Acid), steroids injection, Body Mass Index (BMI), medical procedures (e.g. hip and knee replacement), medical images (e.g. x-ray reports indicating osteoarthritis), referral to orthopedic surgeon, and referral to physiotherapy. In addition, we created two new variables: 1) medication prescriptions, which combined analgesics, NSAIDs or steroids injection 2) Referral, which combined referral to orthopedic surgeon or physiotherapy. Table 2 summarizes the association between those variables and osteoarthritis diagnosis. Each variable, except for steroid injection, met the criteria (OR \ge 2 or p-value \le 0.1).

[Insert Table 2 here]

Multivariate Logistic Regression Model: Seven of the ten potential predictors were found to be statistically significantly associated with osteoarthritis diagnosis. These variables included: billing, condition, analgesics, NSAIDs, procedures, images, and referral to orthopedic surgeon. The final model demonstrated 81% sensitivity, 90% specificity, 87% PPV, 84% NPV, and 89.9% Area Under the Curve (AUC) of the Receiver Operating Characteristic (Tables 3 and 4, Figure 2).

[Insert Table 3 here]

[Insert Table 4 here]

[Insert Figure 1 here]

Validating the model: After splitting the dataset randomly in half, the model was re-created in one half and 5 of the 7 predictors remained significant. These predictors included: condition, analgesics, procedures, images, and referral to orthopedic surgeon. This model of five predictors was then tested in the other half and demonstrated 67% sensitivity, 99% specificity, 98% PPV, 78% NPV, and 82.7% AUC (Table 5, Figure 2).

[Insert Table 5 here]

[Insert Figure 2 here]

Sensitivity Analysis: We tested the significance of the algorithm after excluding two variables that are not currently available through CPCSSN data (procedures, images). This model yielded 72% sensitivity, 86% specificity, 81% PPV, 78% NPV, and 82.5% AUC (Figure 3).

[Insert Figure 3 here]

Discussion

Summary of Main Findings

The existing CPCSSN algorithm for identifying patients with osteoarthritis had perfect specificity (100%), but poor sensitivity (45%). A multivariate logistic regression model was used to identify characteristics that predicted osteoarthritis diagnosis. By adding five variables (analgesics, NSAIDs, procedures, images, and referral) to the existing algorithm, the sensitivity improved from 45% to 81% while still maintaining high specificity (90%). The new algorithm was then re-created in half of the dataset with five of those seven predictors remaining significant. Then, the algorithm was tested in the second half of the dataset, yielding near-perfect specificity (99%) with a relatively high sensitivity (67%). This reduction in sensitivity was expected since the model was developed in a different dataset. The original dataset provides the best possible outcome, whereas the test dataset is more reflective of the estimates that are anticipated in the general population. Therefore, the model appeared to be rigorous since it produced better sensitivity than the existing CPCSSN algorithm.

The new algorithm contained five predictors that are readily available in the CPCSSN data. These included: billing, condition, analgesics, NSAIDs, and referral to orthopedic surgeon. Two variables that are not currently available in CPCSSN data are the procedures and images. A sensitivity analysis in which both of these variables were excluded from the model reflected high sensitivity and specificity (72% and 86%, respectively).

Comparison with Existing Literature

Generally, little research has evaluated diagnostic algorithms for osteoarthritis surveillance in primary care databases. Most of the research into this area has been done using administrative data. The existing CPCSSN algorithm has shown similar accuracy measures, if not superior accuracy, compared to claims-based algorithms. For example, a study that evaluated administrative data algorithms in identifying patients with arthritis had a sensitivity of 75% and a specificity of 57% compared to the CPCSSN algorithm which had 45% sensitivity and 100% specificity (7). Another study which examined an algorithm for identifying people with osteoarthritis using administrative data had 62% PPV and 78% NPV compared to the CPCSSN algorithm which has 100% PPV and 68% NPV (8). Finally, another study that examined the accuracy of several diagnostic algorithms using administrative data reported a range of sensitivity (7% to 55%) and specificity (77% to 99%) (9).

When considering the characteristics or variables that identify patients with osteoarthritis, the model was generally consistent with existing literature. For example, previous studies have developed a diagnostic algorithm that relied on arthritis-related drugs, outpatient and emergency department diagnoses, procedures and lab tests when identifying patients with arthritis (7). All of these variables were found to be significantly associated with osteoarthritis diagnosis and were thus included in the model. However, when these variables were accounted for, the diagnostic algorithm demonstrated superior sensitivity (81%) and specificity (90%) compared to administrative-based algorithms that included the same variables (75% sensitivity and 57% specificity) (7). A large systematic review has concluded that diagnostic algorithms for symptomatic chronic diseases such as osteoarthritis require clinical evidence to achieve optimal accuracy (10). This conclusion is consistent with our finding that when clinical evidence, such as x-ray images, was added to the model, the sensitivity improved from 45% to 81%.

Strengths and Limitations

This study is the first to explore ways to improve the accuracy of the CPCSSN casefinding diagnostic algorithms. Since CPCSSN providers have already agreed to allow audits of medical records, no issue was encountered in accessing data. This demonstrates a significant improvement over previous validation studies that had low response rate. For instance, a validation study using an administrative database in Ontario had a response rate of only 11% (11). A low response rate could lead to selection bias and a lack of generalizability because responding practices are systematically different from non-responding practices (12). In our study, medical records from all randomly selected patient charts were examined.

The sample size was relatively large compared to similar studies done in the United Kingdom Clinical Practice Research Datalink. Also, the random sampling technique utilized in this study improves on previous research that employed disease-specific stratified sampling, which was based on information for cases alone. This type of sampling was limited in that it could not be used to calculate sensitivity and specificity (12). The positive predictive value is dependent on the prevalence of disease, and thus it might not provide a complete picture of the validity of diagnostic algorithms (12). Our study addresses this limitation by employing an age-stratified random sample from all patients. By validating non-cases, this study ensures that

all patients (cases and non-cases) are subject to the same criteria (12). It also avoids verification bias which results from assessing the accuracy of a test on cases only (13).

The limitations of this study ought to be considered before any conclusions can be drawn. One limitation of the study is that the validation of the new algorithm was done internally, using the same data but splitting it randomly in half. This has been used previously (6). However, an alternative would be to test the new algorithm in an external population. We did not have the resources to conduct this type of validation.

Since the data was obtained from one EMR platform, it was not possible to assess whether regional differences and different EMRs had an impact on validity measures. Moreover, the study could not assess the generalizability of findings across the entire CPCSSN database. There are at least two other similar projects being currently conducted in Alberta and Manitoba. Once these findings are known, validity measures can be compared across the different sites. Furthermore, the newly developed algorithm can be tested in those sites to demonstrate external validation.

The chart abstraction was based on the electronic records available through the Kingston EMR platform entitled OSCAR. It is important to consider the level of completeness of these records to understand the limitations associated with this process. OSCAR records including physician notes and medications are available for all patients starting from 2004. Other data including test results, specialist and hospital correspondence were available from June 2010 for all patients. Prior to 2010, documents were imported into OSCAR selectively for some patients based on physicians' requests.

It is also important to acknowledge that the study was subject to measurement error since a true gold standard is not available. Primary chart abstraction is an imperfect gold standard and thus is subject to measurement error (13). However, this gold standard is less problematic than self-reports that are used by validation studies for chronic diseases. Previous evidence has indicated that people with chronic conditions under-report their illnesses (14). Thus, primary chart abstraction based on clinically-verified data can be considered more accurate than self-reported data.

Implications

Although the existing CPCSSN algorithm had perfect specificity (100%), it had poor sensitivity (45%). These validity measures indicate that osteoarthritis patients identified through CPCSSN were true cases. In other words, the current algorithm eliminates false positives and thus is useful for researchers who are interested in selecting a cohort of true osteoarthritis cases. However, the poor sensitivity of the algorithm implies that not all cases were identified. In fact, approximately 55% of the cases were missed. Thus, it can be argued that the osteoarthritis patients who were identified through CPCSSN may not fully represent the overall osteoarthritis case mix. One difference between these two patient populations (cases identified through the gold standard versus those only identified through CPCSSN) is in the severity of the disease. It can be argued that cases identified through CPCSSN are more likely to exhibit more severe osteoarthritis than those cases identified through the gold standard. This observation is supported by evidence from the chart abstraction in which approximately 18% of cases were identified through medical images and physician notes and

17% were identified through physician notes only. In other words, approximately 35% of osteoarthritis patients were not referred to secondary care, did not have hip or knee replacement procedures, and were not listed under the osteoarthritis disease registry by their physicians. This finding indicates that a large proportion of cases did not undertake extensive medical intervention, which could be why many of these cases were missed by the CPCSSN algorithm. This also suggests that osteoarthritis for this group of patients was not a major burden and did not require comprehensive clinical care. Nevertheless, these patients had osteoarthritic joints as evidenced by the x-ray reports (18%) or physician notes (17%). This observation raises questions on how should osteoarthritis be defined within this context. It can be argued that having a robust algorithm that has perfect specificity and can capture the more severe cases of osteoarthritis is justified for the purpose of identifying a highly selective yet not representative cohort of osteoarthritis cases.

If the purpose of the diagnostic algorithm is to identify the majority of osteoarthritis cases in order to plan and allocate necessary resources for management, then a diagnostic algorithm that maximizes sensitivity is necessary. However, an algorithm with high sensitivity yet poor specificity is also problematic because it may not differentiate between osteoarthritis patients and those with related disorders, such as rheumatoid arthritis and gout.

Therefore, it is important to develop an algorithm that maximizes both sensitivity and specificity. We achieved this purpose by using a multivariate logistic regression and Receiver Operating Characteristic curve. The new algorithm consisted of two variables that exist in the current CPCSSN algorithm (billing data and condition) and added five other variables. Three of those additional variables are readily available through CPCSSN data: analgesics, NSAIDs, and referral to orthopedic surgeon. The other two additional variables have been found to be useful in identifying patients with osteoarthritis, but were not currently available through CPCSSN: procedures and images indicating osteoarthritis.

When all seven characteristics to identify people with osteoarthritis were included in the model, the algorithm had better sensitivity and specificity (81% and 90%, respectively). Therefore, it is recommended that an investment is made to ascertain data on these two variables that are not currently available. There are a number of solutions that can be utilized to extract such data. For example, natural language processing has been used to extract data from radiographic images through electronic text screening (15). However, even after the removal of these two variables, the model still maintained relatively high sensitivity and specificity (72% and 86%, respectively).

Conclusions

The current diagnostic algorithm for osteoarthritis had perfect specificity but poor sensitivity. A logistic model for identifying patients with osteoarthritis improved sensitivity and maintained high specificity. This model was validated in a sample dataset. These findings suggest adding three variables to the current CPCSSN diagnostic algorithm, including analgesics, NSAIDs, and referral to orthopedic surgeon. These variables are readily available through CPCSSN data. In addition, we recommend investments that would allow for the extraction of data on medical images and procedures. This approach can be used towards further refining the diagnostic algorithms for other chronic conditions.
References

1. Schoen C, Osborn R, Doty MM, Bishop M, Peugh J, Murukutla N. Toward higherperformance health systems: adults' health care experiences in seven countries, 2007. Health.Aff.(Millwood). 2007 Dec;26(6):w717-34.

2. 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.

3. Lawrence RC, Felson DT, Helmick CG, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58(1):26–35.

4. Birtwhistle R, Keshavjee K, Lambert-Lanning A, et al. Building a Pan-Canadian Primary Care Sentinel Surveillance Network: Initial Development and Moving Forward. The Journal of the American Board of Family Medicine. 2009 Jul;22(4):412-422.

5. Shah BR, Hux JE, Laupacis A, Zinman B, Cauch-Dudek K, Booth GL. Administrative data algorithms can describe ambulatory physician utilization. Health Serv.Res. 2007 Aug;42(4):1783-1796.

6. 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 Sep 15;59(9):1314-1321.

 Harrold LR, Li W, Yood RA, Fuller J, Gurwitz JH. Identification of patients with arthritis and arthritis-related functional limitation using administrative data. J Public Health Manag Pract 2008 Sep-Oct;14(5):487-497.

8. Harrold LR, Yood RA, Andrade SE, Reed JI, Cernieux J, Straus W, et al. Evaluating the predictive value of osteoarthritis diagnoses in an administrative database. Arthritis Rheum 2000 Aug;43(8):1881-1885.

9. Rector TS, Wickstrom SL, Shah M, et al. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. Health Serv Res 2004 Dec;39(6 Pt 1):1839-1857.

10. Raina PS, Wolfson C, Kirkland SA, Keshavarz H, Griffith LE, Patterson C, et al. Ascertainment of chronic diseases in the Canadian longitudinal study on aging (CLSA), systematic review. Can J Aging 2009 Sep;28(3):275-285.

11. 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 Mar;25(3):512-516.

 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 Jan;69(1):4-14.

13. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007 Feb 6;115(5):654-657.

14. Turner, C. F., T. K. Smith, L. K. Fitterman, et al. 1997. "The Quality of Health Data Obtained in a New Survey of Elderly Americans: A Validation Study of the Proposed Medicare

Beneficiary Health Status Registry (MBHSR)." Journal of Gerontology: Social Sciences 52B (1): S49–58.

15. van Walraven C, Wong J, Morant K, Jennings A, Austin PC, Jetty P, et al. Radiographic monitoring of incidental abdominal aortic aneurysms: a retrospective population-based cohort study. Open Med 2011;5(2):e67-76.

16. Kadhim-Saleh A., Green M., Williamson T., Hunter D., Birtwhistle R. Validation of the diagnostic algorithms for five chronic conditions in the Canadian Primary Care Sentinel Surveillance Network. The Journal of the American Board of Family Medicine. Forthcoming 2012.

17. SAS Institute Inc., Version 9.2, Cary, NC.

Tables

Table 1. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) along with the 95% confidence intervals, existing CPCSSN algorithm (billing data, condition) versus chart abstraction.

		Chart A	udit Dx]	Sensitivity	Specificity	PPV	NPV
thritis	CPCSSN Dx	Yes	No		45%	100%	100%	68%
teoar	Yes	65	0	*	0.37-0.54	0.98-1	0.94-1.0	0.62-0.74
Os	No	79	169					

*Confidence Intervals (Exact)

Table 2. Summary of the Univariate Analysis, detailing the association between each variableand osteoarthritis diagnosis based on the gold standard. Chi-square statistic, p-value, OddsRatio and 95% Confidence Interval, as well as sensitivity and specificity are provided.

Variable	Chi-square	p-value	OR (logit)	95% CI	Sensitivity	Specificity
Billing data	30.5	<.0001	68.9	4.2 - 1144.5	17%	100%
Condition	76.6	<.0001	204.1	12.5 - 3344.3	38%	100%
Analgesics	30.3	<.0001	6.5	3.1 - 13.7	29%	94%
NSAIDs	41.1	<.0001	7.2	3.7 - 13.9	38%	92%
Injection	1.5	0.2179	1.6	0.75 - 3.4	12%	92%
BMI	8.1	0.0043	1.9	1.2 - 3.1	65%	51%
Procedures~	41.5	<.0001	51.9	7.0 - 384.9	24%	99%
Images~	87.2	<.0001	127.0	17.3 - 932.3	43%	99%
Orthopedic Surgeon Referral	19.2	<.0001	13.5	3.1 - 58.7	14%	99%
Physiotherapy Referral	7.8	0.0053	3.6	1.4 - 9.5	12%	96%
Medication*	42.2	<.0001	5.0	3.0 - 8.3	55%	80%
Referral**	27.7	<.0001	6.9	3.1 - 15.5	26%	95%

OR=Odds Ratio; CI=Confidence Intervals; NSAIDs= Non-steroidal Anti-Inflammatory Drugs (excluding acetylsalicylic acid and naproxen); BMI= Body Mass Index; *Medication includes prescription of analgesic, NSAIDs or injection; **Referral includes referral to orthopaedic surgeon or physiotherapy; ~data from procedures and images were only available through chart abstraction

Table 3: Building the osteoarthritis diagnostic algorithm in the full dataset.

Characteristic	Adjusted OR	Log OR	P-value
Analgesics	2.5	0.9	0.188
Ortho referral	7.4	2.0	0.067
Billing data	16.9	2.8	0.011
NSAIDs	4.4	1.5	0.003
Procedures*	29.8	3.4	0.0006
Images*	86.7	4.5	<0.0001
Condition	48.4	3.9	<0.0001

Full Dataset (n=313)

OR=Odds Ratio; NSAIDs= Non-steroidal Anti-Inflammatory Drugs; *Data on procedures and images were collected through chart abstraction; log OR indicates the prediction score associated with each variable

Table 4. Results of the final model. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are provided, CPCSSN algorithm (analgesics, ortho referral, NSAIDs, billing data, procedures, images, condition) versus chart abstraction.

		Chart Audit Dx]	Sensitivity	Specificity	PPV	NPV
Osteoarthritis	Meets algorithm cut- off score*	Yes	No		81%	90%	87%	84%
	Yes	116	17					
0	No	28	152					

*Cut-off score was 1.5

Table 5: Testing the osteoarthritis diagnostic algorithm in a sample dataset.

Characteristic	Adjusted OR	Log OR	P-value
Analgesics	4.4	1.5	0.0615
Ortho referral	23.6	3.2	0.0058
Procedures	38.3	3.6	0.0002
Images	103.4	4.6	<0.0001
Condition	59.4	4.1	<0.0001

Test Dataset (n=158)

OR=Odds Ratio; NSAIDs= Non-steroidal Anti-Inflammatory Drugs; *Data on procedures and images were collected through chart abstraction; log OR indicates the prediction score associated with each variable

Figures



Figure 1. Receiver Operating Characteristic curve of the final model

Point labels are values of _sens_ _spec_

The Area Under the Curve for the final model was approximately 89.8%. The numerator and denominator represent the sensitivity and specificity, respectively.





Point labels are values of _sens_ _spec_

The Area Under the Curve for the model in the test dataset was approximately 82.7%. The numerator and denominator represent the sensitivity and specificity, respectively.





ROC plot for disease = score

Point labels are values of _sens _spec_

The Area Under the Curve for the sensitivity analysis (excluding procedures and images) was approximately 82.5%. The numerator and denominator represent the sensitivity and specificity, respectively.

Chapter 4: General Discussion

Summary of Study

This thesis project was a validation study of the Electronic Medical Records-based diagnostic algorithms in the Canadian Primary Care Sentinel Surveillance Network. The aim of the first manuscript was to assess the validity of these diagnostic algorithms for surveillance of five chronic conditions, including: type 2 diabetes, hypertension, osteoarthritis, COPD, and depression. This was done by conducting primary chart abstraction of patients' medical records. The main conclusion of the first manuscript was that the diagnostic algorithms for type 2 diabetes (100% sensitivity, 99% specificity) and hypertension (83% sensitivity, 98% specificity) were accurate. The diagnostic algorithms for the other three conditions (osteoarthritis, COPD, and Depression) had nearly perfect specificity, but poor sensitivity.

The second manuscript sought to explore ways to enhance the accuracy for the diagnostic algorithms, using osteoarthritis as an example. This was done by developing a multivariate logistic regression model to identify characteristics that are associated with osteoarthritis diagnosis. The model improved sensitivity from 45% to 81%, while maintaining high specificity (90%). The new algorithm was then validated by splitting the dataset in half, recreating the model in one half and testing it in the other half. The main conclusion of manuscript 2 is that the new algorithm was shown to enhance the accuracy of osteoarthritis case-finding.

General Strengths and Limitations

This study is the first validation of the CPCSSN database, considering the accuracy of case-finding diagnostic-algorithms and the ascertainment of a model that maximizes sensitivity and specificity. A rigorous methodological approach was based on a random-sampling strategy, comprehensive abstraction of patients' medical records, and full participation rate. In addition, a rigorous statistical analysis strategy was used to develop an algorithm that maximized sensitivity and specificity for osteoarthritis case-finding (1). After creating the model in the full dataset, the data was randomly split in half, re-developing the model in one half and testing it in the other half.

The limitations of this study ought to be considered before any conclusions can be drawn. One challenge is the use of different EMRs by research networks. Different EMRs have different coding structures and variable data extraction. Since data was examined from only one EMR platform, it was not possible to assess whether regional differences and different EMRs have an impact on validity measures. Additionally, the study was unable to assess the generalizability of findings across the CPCSSN database since it examined accuracy at the Kingston network alone. In other words, data from one network may not represent the overall data quality. Another limitation in the study was that the validation of the new algorithm was done internally, using the same data but splitting it randomly in half. This has been used previously (1). However, an alternative would be to test the new algorithm in an external population, although resources were unavailable to conduct this type of validation. There are at least two other similar projects being currently conducted at two sites in Alberta and Manitoba. Once these findings are known, validity measures can be compared across the different sites. In

addition, the newly developed algorithm can be tested in those sites to demonstrate external validation.

The study was also limited to a retrospective analysis, where clinical records have been already documented by different health providers. Thus, the documentation was done inconsistently (2) and a longitudinal assessment could be considered more accurate than a retrospective design. It is also important to acknowledge the limitation that the study is subject to measurement error since a true gold standard is not available. Primary chart abstraction is an imperfect gold standard and thus is subject to measurement error (3). However, this gold standard is less problematic than self-reports that are used by validation studies for chronic diseases. Previous evidence has indicated that people with chronic conditions under-report their illnesses (4). Thus, primary chart abstraction based on clinically-verified data can be considered more accurate than self-reported data. In addition, the inter-rater reliability of chart abstraction was found to be substantial (100% perfect match for osteoarthritis diagnosis and 89.3% overall Kappa Statistic).

Overall Challenges

It is important to outline the general challenges and lessons learned through this validation work on Electronic Medical Records. One challenge that was encountered is the issue of defining cases. For example, the second manuscript discusses how 35% of patients with osteoarthritis were defined as cases although they did not have referrals, procedures, or medications. One potential solution for this problem could be the specification of criteria for who fits the definition of a case. There was a general agreement that patients were identified as cases if the physician notes or other clinical evidence such as x-ray reports confirm. However, the problem of this approach is that the severity of a case could not be ascertained. Another challenge was the issue of identifying a clear purpose for the diagnostic algorithms. Having this purpose explicitly articulated could have implications on the disease definition and how data collection was implemented. For example, if the purpose of the algorithm is to identify every possible case regardless of severity of disease, then our data collection protocol was adequate. However, if the purpose is to identify a selective cohort of patients who utilize health services frequently and would benefit from a management strategy, then a more narrow definition could have been more appropriate. Therefore, one lesson learned through this validation work is to align the definition of a case along with the purpose of the algorithm.

Another challenge was the issue of extracting data from a source that was created for clinical rather than research purposes. Electronic Medical Records have tremendous amount of clinical information, however it is difficult to extract the relevant information in a systematic way. This challenge was addressed by designing the standardized form according to the unique attributes of the EMR platform from the Kingston site. The collection tool was also developed based on consultations with the clinicians who use this platform. Although these steps took considerable time, it was a necessary process to ensure that data collection was done consistently. Therefore, the other lesson learned through this project is to carefully examine the data source (EMR in this case) and consult with individuals who actively use it.

Implications

There are a number of direct and indirect implications for this study. First, the study has demonstrated that the diagnostic algorithms for diabetes and hypertension case-finding were adequate, thus researchers and policy makers can utilize the CPCSSN data reliably. Secondly, the current algorithms for osteoarthritis, COPD and depression have nearly perfect specificity. Thus, these algorithms can help researchers in identifying patients with these conditions with certainty. However, the current algorithms for these conditions have poor sensitivity, thus under-estimating the true prevalence. Based on these findings, we sought to maximize the sensitivity and specificity for these algorithms, using osteoarthritis as an example. The results from the second manuscript demonstrate that a new algorithm for osteoarthritis case-finding improves sensitivity from 45% to 81% and maintains high specificity (90%). This new algorithms includes five additional characteristics that were not initially included in the current algorithm. Three of those characteristics are readily available through CPCSSN data and thus can be included to enhance sensitivity of the model. Two characteristics (medical images and procedures) are not readily available through CPCSSN, but investments can be made to extract this data. For example, natural language processing has been used to extract data from radiographic images through electronic text screening (5). Another implication of this project is that the approach utilized to develop an algorithm that enhances sensitivity and specificity can be applied towards the other two conditions, COPD and depression.

This thesis project directly informs our understanding of the CPCSSN data quality at the Kingston site. However, the project also informs researchers at other sites regarding the validation process. Finally, this project informs researchers of the validation of other chronic

conditions, specifically Parkinson's, Dementia, and Epilepsy. This work has already been started at the Kingston and Alberta practice-based research networks.

Conclusions

The accuracy of EMR-based diagnostic algorithms for two chronic conditions, diabetes and hypertension, have been shown to be adequate. This finding ensures that researchers and policy-makers can utilize these algorithms reliably. The algorithms for the other three conditions (osteoarthritis, COPD, depression) demonstrate near-perfect specificity, thus can be used reliably to identify patients with those conditions. There is little research conducting on ways to improve the accuracy of diagnostic algorithms in EMR-based information systems. This thesis contributes to this area of research by developing a model that identifies patients with osteoarthritis. This model demonstrated better accuracy compared to the current algorithm. Therefore, the findings of this thesis have implications directly relevant to the CPCSSN database, but also can be used to further our understanding of data quality of EMR-based information systems. Future studies on this area should explore ways to improve the accuracy for other diagnostic algorithms, such as COPD and depression.

References

- 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 Sep 15;59(9):1314-1321.
- Harrold LR, Yood RA, Andrade SE, Reed JI, Cernieux J, Straus W, et al. Evaluating the predictive value of osteoarthritis diagnoses in an administrative database. Arthritis Rheum 2000 Aug;43(8):1881-1885.
- 3) Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007 Feb 6;115(5):654-657.
- 4) Turner, C. F., T. K. Smith, L. K. Fitterman, et al. 1997. The Quality of Health Data Obtained in a New Survey of Elderly Americans: A Validation Study of the Proposed Medicare Beneficiary Health Status Registry (MBHSR). Journal of Gerontology: Social Sciences 52B (1): S49–58.
- 5) van Walraven C, Wong J, Morant K, Jennings A, Austin PC, Jetty P, et al. Radiographic monitoring of incidental abdominal aortic aneurysms: a retrospective population-based cohort study. Open Med 2011;5(2):e67-76.

Appendices

Diabetes	Depression	Osteoarthritis	Hypertension	COPD
1. Billing ICD9- 250x; Service code in Ontario K030A or Q040A, in other provinces consult physician and document the code atleast twice in the last 2 years	1. Billing codes 296, 311	1. Billing ICD0- 715x, 721	1. Billing ICD 9-401 to 405 at least twice in the last two years	1. Billing ICD9- 492, 496
2. In problem list Diabetes/NIDDM / DM	 In problem list Depression, post traumatic stress disorder, 311 	2. In problem list osteoarthritis	2. In problem list Hypertension, HTN	2. In Problem list Bronchitis, Emphysema, COPD/COLD
 3. In the medication list: Insulin; sulfonylureas:Tol butamide, Glibenclamide,Gl iclazide, Glimiperide; biguanides Metformin; Others: Acarbose, Sitagliptin 	3. In Medication list SSRI: fluoxetine, Fluvoxamine, Setrtaline; MAOI: MoclobemideTr anylcypromine; Tricyclics (>75mg/d): Imipramine, Amitrityptiline; Doxepin,	3.In medication list : none identified	 3. In mediation list ACEi/ARB: Captopril,Enalapril,Lisi nopril,cilazapril, Benazepril; Diuretics: Hydrochlorthiazide,Chl orthalidone, amiloride, Trianterine, metalozone; Beta blockers: Timolol, acebutalol, atenolol, metoprolol; Calcium Channel blockers: 	3. In medication list: Beta agonists:Salbut amol, salmetrol, Formoterol Anticholinergics :ipratropium ; Xanthines: ;, Inhalant corticosteroid:fl uticasone

Appendix 1. Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Disease Definitions

	desipramine,		Verapamil, diltiazem,	
	clomipramine,		Filodipine,Amlodipine;	
	maprotiline,		Alpha blockers:	
	Timipramine		Methyldopa; others:	
			Hydralazine, Clonidine,	
			Indapamide	
4. In the lab				4. Risk factors
result: HbA1C				Smoker,
>7% at anytime,				age>=35 years
Fasting BS >7				
twice in the last				
1 year				
EXCLUSIONS:	EXCLUSIONS:	EXCLUSIONS:	EXCLUSIONS:	EXCLUSIONS:
Polycystic	Anxiety		Migraines 346, CHF	Asthma 493
Polycystic ovarian	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial	Asthma 493
Polycystic ovarian syndrome 256.4,	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412,	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8,	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1.	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidnev	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249.	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249, Hypoglycemia	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal hypertension 572.3	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249, Hypoglycemia NOS 790.29	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal hypertension 572.3	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249, Hypoglycemia NOS 790.29, neonatal	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal hypertension 572.3	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249, Hypoglycemia NOS 790.29, neonatal diabetes mellitus	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal hypertension 572.3	Asthma 493
Polycystic ovarian syndrome 256.4, Gestational diabetes 648.8, Secondary chemical induced diabetes 249, Hypoglycemia NOS 790.29, neonatal diabetes mellitus 775 1	Anxiety disorders 300		Migraines 346, CHF 428, Myocardial infarction 410 &412, diabetes 250, cardiac arrythmia 427, Tremor 333.1, Esophageal varices 456.0 &456.1, angina 413, Kidney stones 592, Portal hypertension 572.3	Asthma 493

Appendix 2: Chart Abstraction Form

"A validation study of computer-based diagnostic algorithms for chronic disease surveillance"

CHART ABSTRACTION

How to Complete This Chart Abstraction

The chart abstraction is divided into 3 sections and deals with patients' diagnoses. Please ensure the confidentiality of the patients and providers and do not record other additional personal information on these forms. Record the Patient CPCSSN ID and the Practice Site ID #, in addition to the date that the abstraction was conducted. Please follow any instructions or advice that is offered by the staff at the practice site. **SECTION 1: Context**

1. Year of birth:_____ Age : _____

2. Sex: 🛛 Male 🖓 Female

SECTION 2: Data extraction

A. EMR problem list/disease registry

A.1. Is there a diagnosis of diabetes in the EMR problem list/disease registry (<u>Diabetes/NIDDM/DM</u>)?

□ No □ Yes

A.2. Is there a diabetes flow sheet?

🗆 No 🛛 🖾 Yes

A.3. Is there a diagnosis of depression in the EMR problem list/disease registry (<u>Depression, post-traumatic stress disorder</u>)?

□ No □ Yes

A.4. Is there "depression" under past medical history?

□ No □ Yes

A.5. Is there a diagnosis of hypertension in the EMR problem list/disease registry (<u>Hypertension</u>, <u>HTN</u>)?

🗆 No 🛛 🖾 Yes

A.6. Is there a hypertension flow sheet?

□ No □ Yes

A.7. Is there a diagnosis of osteoarthritis in the EMR problem list/disease registry (Osteoarthritis)?

□ No □ Yes

A.8. Is there a diagnosis of COPD in the EMR problem list/disease registry (<u>Chronic Bronchitis</u>, <u>Emphysema</u>, <u>COPD/COLD</u>)?

□ No □ Yes

B. Is there any of the following under medication list?

1.	Diabetes:			Record whether any of the medications were prescribed for other conditions (POS,		
	B.1.1. Insulin	🗆 No 🗖 Y	es	Gestational DM, Secondary DM, Hyperglycemia NOS, Neonatal DM)		
	B.1.2. Glyburide (Sulfonylurease)	🗆 No 🗖 Ye	25			
	B.1.3. Metformin (Biguanides)	🗆 No 🗖 Ye	25			
2.	Depression:			Record whether any of the		
	B.2.1. SSRI	🗆 No 🔲 Yes	Date:	medications were prescribed for other conditions (anxiety)		
	B.2.2. MAOI	🗆 No 🖾 Yes	Date:			
	B.2.3. Tricyclics (>75 mg/day)	🗆 No 🗖 Yes	Date:			
3.	Hypertension:					
	B.3.1. ACEI/ARB	🗆 No 🗖 Ye	25	Record whether any of the medications were prescribed for other conditions (Migraines		
	B.3.2. Diuretics	🗆 No 🖾 Ye	25	CHF, Myocardial Infarction, Diabetes, Cardiac Arrhythmia, Tremor, Esophageal Varices,		
	B.3.3. Beta Blockers	🗆 No 🗖 Ye	25	Hypertension)		
	B.3.4. Calcium Channel Blockers	🗆 No 🗖 Ye	25			
	B.3.5. Alpha Blockers	No Ye	25			

4. COPD

B.4.1. Spiriva (Tiotropium bromide)	🗖 No	🖵 Yes	Record whether any of the medications were prescribed for other conditions (asthma)
B.4.2. Beta agonists	🗖 No	🖵 Yes	
B.4.3. Anticholinergics Xanthines	🖵 No	☐ Yes	
B.4.4. Inhalant corticosteroids	🗖 No	🖵 Yes	

C. EMR_ Physiological (record all the following):

Weight

BMI_____

Height

Record up to the 3 most recent <u>in-</u> <u>office</u> Blood Pressure readings in the past 2 years [from three different visits]	Date Done MM/DD/YY	Taken/ No Results	Recommended & Refused	Not Done
1st. Bp:mmHg	//	🗆 NR	🗖 RR	🗖 ND
2nd. Bp:mmHg	//	🗆 NR	🗖 RR	🗖 ND
3rd. Bp:mmHg	//	🗖 NR	🗖 RR	🗖 ND

D. Under lab test results:

1. Diabetes

	D.1.1.	HbA1c testing done	🗖 No	🖵 Yes		
	D.1.2. D.1.3. D.1.4.	If yes, was HbA1c ≥ 7% Date Recorded (YY/MM/DD) Exact value	□ No	🛛 Yes		
	D.1.5.	Fasting Blood Sugar testing done	🗖 No	🛛 Yes		
	D.1.6.	If yes, was Fasting BS > 7	🛛 No	🛛 Yes		
	D.1.7.	Date Recorded (YY/MM/DD)				
	D.1.8.	Exact value				
2.	Depression D.2.1.	Any of the following tests done: (PHQ-9), HAMD-T	7, Geria	tric Depr D No	ession scale) □ Yes
	D.2.2.	If yes, was there diagnosis of depression	n		🗖 No	Tes Yes
3.	Osteoarthri	itis				
	D.3.1. osteoarthri	Any X-ray and other joint imaging (e.g. tis investigation	MRI) repo	orts don	e relatec D No	l to Types
	D.3.2.	If yes, was there diagnosis of osteoarth	ritis		🛛 No	Tes Yes

4. COPD

D.4.1. reports)	Any of the following done (Pulmonary Function Tests, C	Chest Xray	//Chest CT
D.4.2.	If yes, was there diagnosis of COPD	🛛 No	🖵 Yes

E. Referrals:

E.1.1. Diabetes-related referral (ophthalmology, nephrology, neurology, endocrinology, Diabetes education, dietician). Only record if referral was related to diabetes.

🗆 No 🛛 🖾 Yes

E.1.2. If yes, which one(s): _____

E.2.1. Depression-related referral (psychiatrist, psychologist) Only record if referral was related to depression:

🗆 No 🛛 🖾 Yes

E.2.2. If yes, which one(s): _____

E.3.1. Osteoarthritis-related referral (Orthopedic surgeon, Physiotherapy, Occupational therapy, Exercise program). Only record if referral was related to osteoarthritis:

□ No □ Yes

E.3.2. If yes, which one(s): ______

E.4.1. COPD-related referral (Respirology, Thoracic surgery). Only record if referral was related to COPD:

□ No □ Yes

E.4.2. If yes, which one(s): ______

F. Procedures:

F.1.1. Diabetes-related procedures (Revascularization procedure, Coronory artery disease investigation):

□ No □ Yes

F.1.2. If yes, which one(s): ______

F.2.1. Osteoarthritis-related procedures (Knee replacement, Hip replacement):

🗆 No 🛛 🖾 Yes

F.2.2. If yes, which one(s): ______

F.3.1. COPD-related procedures (Spirometry):

🗆 No 🛛 🖾 Yes

F.3.2. If yes, which one(s): ______

G. Hospitalizations:

G.1. Hospitalization (Discharge with depression diagnosis)
Q No Q Yes

H. Billing data:

H.1. Diabetes-related codes (250.x)	🗆 No 📮 Yes
H.2. Depression-related codes (309.0, 309.1, 309.8	1; OR 296, 300, 309, 311, 648)
	🗆 No 📮 Yes
H.3. Hypertension-related codes (401-405)	🗆 No 📮 Yes
H.4. Osteoarthritis-related codes (715.x)	🗆 No 📮 Yes
H.5. COPD-related codes (490-492, 496)	🗅 No 🗳 Yes

SECTION 3: diagnoses

1. Diabetes

Summary of section 2 (data extraction)

- 3.1.1.Problem list/disease registry \Box
- 3.1.2.Diabetes flow sheet \Box
- 3.1.3.HbA1c ≥ 7% 🗖
- 3.1.4.Fasting blood sugar >7 \Box
- 3.1.5.Insulin 🗖
- 3.1.6.0 ther medications (e.g. metformin) \Box

<u>Section</u>

Does information contra-indicate or confirm diagnosis?

Lab test results 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Referral 🗖	Contra-indicate 🛛	Confirm 🖵 Explain
Procedures 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Risks 🗖	Contra-indicate 🛛	Confirm 🖵 Explain
Physician notes 🗖	Contra-indicate 🛛	Confirm 🖵 Explain
Hospital correspondence 🛛	Contra-indicate 🛛	Confirm 🖵 Explain
Specialist correspondence 🖵	Contra-indicate 🗖	Confirm 🖵 Explain
Free-text fields 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Billing data 🗖	Contra-indicate 🗖	Confirm 🖵 Explain

According to the chart abstraction, is there evidence that this patient has been diagnosed with diabetes?

Confirmed positive
Unsure (Untested)
Confirmed negative
Confirmed negative

Please comment on any information found in free-text, hospital and specialist correspondence, physicians' notes that influenced your conclusion. Also, comment on whether there is a discrepancy between the default diagnosis and your conclusion:

2. Depression

Summary of section 2 (data extraction)

3.2.1. Problem list/disease registry 🗖

3.2.2. Medications 🖵

Section

3.2.3. EMR referral (psychiatrist, psychologist) 🖵

3.2.4. Hospitalization (discharge with depression diagnosis) \Box

Does information contra-indicate or confirm diagnosis?

Lab test results 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Physician notes 🖵	Contra-indicate 🛛	Confirm 🖵 Explain
Hospital correspondence 🗖	Contra-indicate 🛛	Confirm 🖵 Explain
Specialist correspondence 🖵	Contra-indicate 🗖	Confirm 🖵 Explain
Free-text fields	Contra-indicate 🗖	Confirm 🖵 Explain
Billing data 🖵	Contra-indicate 🗖	Confirm 🖵 Explain

According to the chart abstraction, is there evidence that this patient has been diagnosed with depression?

Confirmed positive

Unsure (Untested) 🗖

Confirmed negative \Box

Please comment on any information found in free-text, hospital and specialist correspondence, physicians' notes that influenced your conclusion. Also, comment on whether there is a discrepancy between the default diagnosis and your conclusion:

3. HYPERTENSION

Summary of section 2 (data extraction)

3.3.1. Problem list/disease registry

- 3.3.2. Hypertension flow sheet \Box
- 3.3.3. Blood Pressure >140/90 on more than 1 reading 🖵
 - For diabetes patients only, Blood Pressure > 140/80 on more than 1 reading
- 3.3.4. Medications 🖵

Section

Does information contra-indicate or confirm diagnosis?

Physiologic Data 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Physician notes 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Hospital correspondence \Box	Contra-indicate 🗖	Confirm 🖵 Explain
Specialist correspondence \Box	Contra-indicate 🗖	Confirm 🖵 Explain
Free-text fields 🖵	Contra-indicate 🗖	Confirm 🖵 Explain
Billing data 🖵	Contra-indicate 🗖	Confirm 🖵 Explain

According to the chart abstraction, is there evidence that this patient has been diagnosed with hypertension?

Confirmed positive

Unsure (Untested) 🗖

Confirmed negative

Please comment on any information found in free-text, hospital and specialist correspondence, physicians' notes that influenced your conclusion. Also, comment on whether there is a discrepancy between the default diagnosis and your conclusion:

4. OSTEOARTHRITIS

Summary of section 2 (data extraction)

3.4.1. Problem list/disease registry 🗖

3.4.2. Referrals 🗖

3.4.3. Procedures 🖵

<u>Section</u>	Does information contra-indicate or confirm diagnosis?	
Lab test results 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Risks 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Physician notes 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Hospital correspondence 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Specialist correspondence	Contra-indicate 🗖	Confirm 🖵 Explain
Free-text fields 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Billing data 🗖	Contra-indicate 🗖	Confirm 🖵 Explain

According to the chart abstraction, is there evidence that this patient has been diagnosed with osteoarthritis?

Confirmed positive
Unsure (Untested)
Confirmed negative
Confirmed negative
Please comment on any information found in free-text, hospital and specialist correspondence, physicians' notes that influenced your conclusion. Also, comment on whether there is a discrepancy between the default diagnosis and your conclusion:

5. COPD

Summary of section 2 (data extraction)

3.5.1. Problem list/disease registry

3.5.2. Spiriva 🗖

3.5.3. Other medications \Box

3.5.4. Referrals 🖵

3.5.5. Procedures 🖵

Section	Does information contra-indicate or confirm diagnosis?	
Risks 🗖	Contra-indicate 🗖	Confirm 🖵 Explain
Physician notes 🗖	Contra-indicate 🖵	Confirm 🖵 Explain
Hospital correspondence 🖵	Contra-indicate 🖵	Confirm 🖵 Explain
Specialist correspondence \Box	Contra-indicate 🖵	Confirm 🖵 Explain
Free-text fields 🖵	Contra-indicate 🖵	Confirm 🖵 Explain
Billing data 🖵	Contra-indicate 🗖	Confirm 🖵 Explain

According to the chart abstraction, is there evidence that this patient has been diagnosed with COPD?

Confirmed positive

Unsure (Untested) 🗖

Confirmed negative **D**

Please comment on any information found in free-text, hospital and specialist correspondence, physicians' notes that influenced your conclusion. Also, comment on whether there is a discrepancy between the default diagnosis and your conclusion:

Appendix 3: Research Ethics Board Approval Letter



OUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD-DELEGATED REVIEW

September 12, 2011

Mr. Amjed Kadhim-Saleh Department of Community Health and Epidemiology Queen's University

Dear Mr. Kadhim-Saleh

Study Title: EPID-357-11 A Validation Study of Computer-Based Diagnostic Algorithms for Chronic Disease Surveillance File # 6006229 Co-Investigators: Dr. M. Green, Dr. D. Hunter, Mr. T. Williamson

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol (2011-08-23) and data collection tool for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following listing of ethics requirements you must fulfill over the course of your study:

Reporting of Amendments: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. Please use event form: HSREB Multi-Use Amendment/Full Board Renewal Form associated with your post review file # 6006229 in your Researcher Portal (https://eservices.gueensu.ca/romeo_researcher/)

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information. Serious Adverse Event forms are located with your post-review file 6006229 in your Researcher Portal (https://eservices.queensu.ca/romeo_researcher/)

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

Albert J. Clark.

Chair, Research Ethics Board September 12, 2011

Investigators please note that if your trial is registered by the sponsor, you must tale responsibility to ensure that the registration information is accurate and complete



QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards as defined by the Tri-Council Policy Statement; Part C Division 5 of the Food and Drug Regulations, OHRP, and U.S DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number: #FWA00004184, #IRB00001173

Current 2011 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board:

Dr. A.F. Clark, Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)

- Dr. H. Abdollah, Professor, Department of Medicine, Queen's University
- Dr. R. Brison, Professor, Department of Emergency Medicine, Queen's University

Dr. M. Evans, Community Member

Dr. S. Horgan, Manager, Program Evaluation & Health Services Development, Geriatric Psychiatry Service, Providence Care, Mental Health Services, Assistant Professor, Department of Psychiatry

Dr. B. S. Kisilevsky, Professor, School of Nursing, Department of Psychology and Obstetrics & Gynaecology, Queen's University

Ms. D. Morales, Community Member

Dr. W. Racz, Emeritus Professor, Department of Pharmacology & Toxicology, Queen's University

Ms. S. Rohland, Privacy Officer, ICES-Queen's Health Services Research Facility, Research Associate, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute Â

Dr. B. Simchison, Assistant Professor, Department of Anesthesiology, Queen's University

Dr. A.N. Singh, WHO Professor in Psychosomatic Medicine and Psychopharmacology Professor of Psychiatry and Pharmacology, Chair and Head, Division of Psychopharmacology, Queen's University, Director & Chief of Psychiatry, Academic Unit, Quinte Health Care, Belleville General Hospital

Dr. E. Tsai, Associate Professor, Department of Paediatrics and Office of Bioethics, Queen's University

Rev. J. Warren, Community Member