CMAJ OPEN
MEDICAL RESEARCH FOR ALL

Manual Review of Electronic Medical Records as a Reference Standard for Case Definition Development

Journal:	*CMAJ Open*
Manuscript ID	CMAJOpen-2017-0077
Manuscript Type:	Cohort (retrospective)
Date Submitted by the Author:	16-Jun-2017
Complete List of Authors:	Williamson, Tyler; University of Calgary, Community Health Sciences Miyagishima, Rebecca; University of Alberta, School of Public Health Derochie, Janeen; University of Calgary, Community Health Sciences Drummond, Neil; University of Alberta, Family Medicine; University of Calgary, Family Medicine
Keywords:	Epidemiology, Family medicine, general practice, primary care, Medical informatics, Statistics and research methods
More Detailed Keywords:	validation studies
Abstract:	Background: The adoption of electronic medical records (EMR) in Canadian primary care provides a valuable opportunity for research and surveillance. To facilitate rigor in the surveillance of chronic disease in Canada, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) has previously undertaken a validation study of case definitions using direct review of 'raw' EMR data. While effective, this method is time-consuming and can present methodological and organizational challenges. We aimed to determine whether the processed and standardized data contained with the CPCSSN database might function as a reference standard for case definition validation. Methods: We compared the case identification results of the chart reviews for eight chronic diseases in 1906 patients with the results of a manual review of the CPCSSN processed data for the same conditions in the same patient sample. Results: A manual review of the CPCSSN records for case ascertainment yielded sensitivity ranging from 77.5% (depression) to 97.2% (diabetes), while specificity was high for all definitions ranging from 92.6% (COPD) to 99.4% (parkinsonism). Positive and negative predictive values (PPV and NPV) demonstrated high accuracy of the manual CPCSSN record review relative to review of the raw chart data. PPV ranged from 83.3% (COPD) to

	93.3% (hypertension) and NPV ranged from 92.4% (osteoarthritis) to 99.7% (epilepsy). Interpretation: The use of CPCSSN records as the reference standard to validate case definitions significantly reduces the burden on sentinel physicians and clinic managers, as well as on researchers, while offering a reference standard that is a reasonable substitution for chart review.

Manual Review of Electronic Medical Records as a Reference Standard for Case Definition Development

Williamson T, Miyagishima RC, Derochie JD, Drummond N

Abstract

Background: The adoption of electronic medical records (EMR) in Canadian primary care provides a valuable opportunity for research and surveillance. To facilitate rigor in the surveillance of chronic disease in Canada, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) has previously undertaken a validation study of case definitions using direct review of 'raw' EMR data. While effective, this method is time-consuming and can present methodological and organizational challenges. We aimed to determine whether the processed and standardized data contained with the CPCSSN database might function as a reference standard for case definition validation.

Methods: We compared the case identification results of the chart reviews for eight chronic diseases in 1906 patients with the results of a manual review of the CPCSSN processed data for the same conditions in the same patient sample.

Results: A manual review of the CPCSSN records for case ascertainment yielded sensitivity ranging from 77.5% (depression) to 97.2% (diabetes), while specificity was high for all definitions ranging from 92.6% (COPD) to 99.4% (parkinsonism). Positive and negative predictive values (PPV and NPV) demonstrated high accuracy of the manual CPCSSN record review relative to review of the raw chart data. PPV ranged from 83.3% (COPD) to 93.3% (hypertension) and NPV ranged from 92.4% (osteoarthritis) to 99.7% (epilepsy).

Interpretation: The use of CPCSSN records as the reference standard to validate case definitions significantly reduces the burden on sentinel physicians and clinic managers, as well as on researchers, while offering a reference standard that is a reasonable substitution for chart review.

Introduction

The adoption of electronic medical records (EMRs) in Canadian primary care practices provides a valuable opportunity to develop research and surveillance-related information[1]. The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) is Canada's only pan-Canadian primary care EMR database. CPCSSN currently holds de-identified records on 1.5M Canadian primary care patients from 1200 sentinel family physicians, nurse practitioners and community pediatricians, in 11 provinces or territories and from 10 different EMR systems. To facilitate rigor in the surveillance of chronic disease in Canada, CPCSSN has previously undertaken a large-scale validation study of case definitions for eight chronic diseases (diabetes, hypertension, osteoarthritis, depression, dementia, chronic obstructive pulmonary disease (COPD), parkinsonism, and epilepsy)[2], which were implemented in the CPCSSN database using computerized algorithms which extract, clean and process the data into a standard format. The case definitions were validated using the accepted reference standard method of manual review of the patient's source electronic medical record. Validation results were favorable for all case definitions, with sensitivity ranging from 77.8 to 98.8, specificity from 93.5 to 99.0, PPV from 72.1 to 92.9 and NPV from 90.2 to 99.9.

While effective, this method of validation is time-consuming and can present challenges to researchers. Access to patient charts must be coordinated with participating clinics, requiring increased time commitment and workload from clinic administrators. Further, there is potential risk to patient privacy and data security which requires additional safeguards to be implemented both by researchers and clinic administrative staff. Thus, it is practical and reasonable to explore alternative sources of reference standard information for use in validation studies. The utility of clinical databases as a reference standard depends upon the effectiveness of the database in representing the information originally contained in the patient charts, as well as the scope of the data available from those charts, given technological and legal constraints. The usefulness of a database as a reference standard may also be condition-specific. A disease may be easily examined within one clinical database but remain obscure for even basic information in another. We aimed

to determine whether the data contained with the CPCSSN records might function as a reference standard for case definition validation.

Methods

In this cross-sectional case definition validation study, we employed the same sample of 1906 patients used in our initial validation study. At the time of that study, the CPCSSN database housed data for 600,000 patients extracted from the EMRs of 475 sentinels. Six out of the ten networks within CPCSSN contributed to the patient sample, which was random and age-stratified such that 90% of the charts were for patients aged sixty years and older. Due to low prevalences, the sample was enhanced by 25 additional charts each for parkinsonism and epilepsy, chosen non-randomly. We compared the case identification results of the record reviews for the eight chronic diseases in the 1906 patients with the results of an independent manual review of the CPCSSN processed data for the same conditions in the same 1906 patients. Note that while the original validation study reported on 1920 patients, 14 were removed from the dataset for this analysis as their clinic is no longer participating in CPCSSN.

CPCSSN Record Review

Two experienced research assistants, one a nurse and the other an epidemiologist, were trained to review patient data records within the CPCSSN database and to assess caseness in each record separately for all 1906 patients, for each of the eight conditions of interest. Each reviewer received a detailed training document which provided instruction on how to assess each record for evidence that the patient should be judged as being a case for each disease. The reviewers were blinded to other assessments of caseness, including the case assignment by CPCSSN's algorithms and the case determination made by reviewers during the original validation study. Reviewers were instructed to examine all aspects of the patient's CPCSSN record to find evidence for caseness, including the list of health conditions, encounter information, medication list, laboratory results, and billing data. Additionally, reviewers could see both the original and cleaned text entries as well as ICD-9

codes. While the review was undertaken independently by the two reviewers, patients for which there was uncertainty were discussed with the team's lead (TW) until consensus was reached. If uncertainty remained, a family physician was consulted for guidance on how to classify the record in question. This approach to resolving discrepancies, including the person to whom the discrepancies were brought, was successfully used in the original study.

Statistical Analysis

Measures of validity employed in this study were sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). These were calculated by comparing the outcomes of the manual CPCSSN record review with the outcomes of the original manual review of the EMR charts. In accordance with the methodology used in Williamson et al. (2014), 70% was considered the cut-off for validity for all measures for sensitivity and specificity. No cut-off value was assigned for PPV or NPV. Additionally, patient demographic information including age, sex and site (that is, the clinic from which a patient's data was extracted) were assessed. All data were analyzed using Stata/IC (Version 13) statistical software.[3]

Results

The analysis was undertaken for all 1906 patients. Table 1 summarizes the patient characteristics of the sample. The sex of patients included in the study was reflective of the intention to over-sample older patients, with 44.3% males and 55.7% females in the final sample. The age of patients ranged from 5 years to 107 years, with 85.5% being over 60 years of age. More than half the patients (50.1%) had a diagnosis noted in the chart according to the chart review. At the time of chart review (2012) these patients had many chronic conditions, with only 22.6% of the sample having none of the 8 conditions under study.

Table 2 outlines the results of the validation analysis using the outcome of chart review as the reference standard. A manual review of the CPCSSN records for case ascertainment

yielded sensitivity ranging from 77.5% (depression, 95% CI: 73.3% to 81.6%) to 97.2% (diabetes, 95% CI: 95.4% to 99.0%), while specificity was high for all definitions and ranged from 93.1% (hypertension, 95% CI: 94.5% to 97.1%) to 99.4% (parkinsonism, 95% CI: 99.0% to 99.8%). PPV and NPV demonstrated high accuracy of the manual CPCSSN record review relative to review of the chart data. PPV ranged from 83.3% (COPD, 95% CI: 77.4% to 89.3%) to 93.3% (hypertension, 95% CI: 91.7% to 94.8%) and NPV ranged from 92.4% (osteoarthritis, 95% CI: 90.9% to 93.8%) to 99.7% (epilepsy, 95% CI: 99.4% to 99.9%). Overall, the case definition for diabetes achieved the highest sensitivity and specificity (97.2% sensitivity, 95% CI: 95.4% to 99.0%; 97.9% specificity, 95% CI: 97.2% to 98.6%).

Discussion

The results of this study demonstrate that CPCSSN record data may function effectively as a reference standard for defining caseness. Agreement in case classification between reviews of CPCSSN records and those of EMR charts was strongest for conditions with the clearest diagnostic criteria (e.g. diabetes) while conditions with less clear diagnostic rules (e.g. depression) showed the largest, but still tolerable, discrepancy.

The use of clinical databases as a source for reference standard data in case definition validation is an expanding topic in primary care epidemiology. Valkhoff et al. (2014) examined billing code and free-text diagnoses of upper gastrointestinal bleeding in two primary care databases (as well as two administrative databases) based in the Netherlands (IPCI) and Italy (HSD).[4] Positive predictive values ranged from 21% (18, 26) for IPCI to 78% (72, 83) for HSD. John et al. (2016) validated Read code validity for anxiety and depression in a Welsh primary care health record database linked with survey results from a community health inequality survey. The authors reported insufficient validity with sensitivity ranging from 0.05 to 0.49[5]. The EMRALD database in Ontario, Canada, based on a single primary care EMR system, has been employed as a reference standard for several case definition validation studies[6-9], however these studies tend to validate case definitions using administrative data with EMR data as the reference standard.

Our interest was to validate the CPCSSN record itself as a possible reference standard against the standard which by convention is generally considered to be preeminent: the medical chart. Little previous work has been reported which sought to rigorously validate case definitions developed using processed EMR data in comparison to those achieved using the conventional reference standard, particularly when linkage to non-primary care data is not feasible. As such, this study is a significant contribution to both primary care and health information technology research in the Canadian context. This has major significance for the development of future case definitions as this will allow researchers to streamline the work and dramatically reduce time and cost constraints which previously presented challenges. This work should lead to significant increases in the number of conditions in the CPCSSN data with validated case definitions, yielding substantial improvement in the utility of the data for research, surveillance and quality improvement studies.

There are limitations to using an EMR database as a reference standard. EMR-derived data is subject to the levels of completeness and accuracy of recording by the entering physician. Missing or erroneous data entered at the clinic-level cannot be addressed by CPCSSN cleaning or coding processes, nor by researchers utilizing the data[10]. Several records in our study had to be excluded due to missing data. However, this will be a similar problem when using chart review as the reference standard. Another limitation relates to the types of data extracted by CPCSSN from the charts. SOAP notes, referral letters, and diagnostic images are among those not extracted from the source EMR record for reasons of confidentiality. If the deterministic information is expected to be found in that type of data then a manual review of the CPCSSN data will not serve well as a reference standard.

Conclusions

This study has demonstrated that chart reviews, which are often challenging to researchers due to time and financial constraints, are a sufficient but sometimes unnecessary reference standard. The use of CPCSSN record data to validate case definitions significantly reduces the burden on sentinel physicians and clinic managers, as well as the researchers themselves. This shorter, more cost-effective process for case definition validation will

increase the potential for future case definition validation work for a variety of conditions occurring in primary care settings.

References

1. Birtwhistle R, Williamson T. Primary care electronic medical records: a new data source for research in Canada. CMAJ 2014; 187(4):239-240.
2. Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, Natarajan N, Manca D, Drummond N. Validating the 8 CPCSSN Case Definitions for Chronic Disease Surveillance in a Primary Care Database of Electronic Health Records. Ann Fam Med 2014; 12(4):367-372.
3. Valkhoff VE, Coloma PM, Masclee GMC, Gini R, Innocenti F, Lapi F, Molokhia M, Mosseveld M, Nielsson MS, Schuemie M, Thiessard F, van der Lei J, Sturkenboom MCJM, Trifirò, on behalf of the EU-ADR Consortium. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. Journal of Clinical Epidemiology 2014; 67:921-931.
4. StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.
5. John A, McGregor J, Fone D, Dunstan F, Cornish R, Lyons RA, Lloyd KR. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. BMC Medical Informatics and Decision Making 2016; 16:35.
6. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, Jaakkimainen L, Lam R, Guttman M. A Validation Study of Administrative Data Algorithms to Identify Patients with Parkinsonism with Prevalence and Incidence Trends. Neuroepidemiology 2014; 43(1):28-37.

7. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, Ivers N, Butt AD, Jaakkimainen RL, Thorne JC, Tu K. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC Musculoskeletal Disorders 2014; 15:216.
8. Schwartz KL, Tu K, Wing L, Campitelli MA, Crowcroft NS, Deeks SL, Wilson SE, Wilson K, Gemmill I, Kwong JC. Validation of infant immunization billing codes in administrative data. Human Vaccines & Immunotherapeutics 2015; 11(7):1840-1847.
9. Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, Jaakkimainen L, Kapral MK. Validity of Administrative Data for Identifying Patients Who Have Had a Stroke or Transient Ischemic Attack Using EMRALD as a Reference Standard. Canadian Journal of Cardiology 2013; 29:1388-1394.
10. Birtwhistle R, Keshavjee K, Lambert-Lanning A, Godwin M, Greiver M, Manca D, Lagacé C. Building a Pan-Canadian Primary Care Sentinel Surveillance Network: Initial Development and Moving Forward. JABFM 2009; 22(4):412-422.

Table 1: Patient Characteristics

Patient Characteristics		Percent
	Male	44.3%
	Age >= 60 years	85.5%
Disease Prevalence		
	Hypertension	50.1%
	Diabetes	16.8%
	Depression	20.3%
	COPD	7.9%
	Osteoarthritis	31.9%
	Dementia	4.9%
	Epilepsy	7.1%
	Parkinsonism	4.4%
Number of chronic conditions (of the 8 CPCSSN conditions)		
	0	22.6%
	1	33.6%
	2	27.1%
	3	12.7%
	4 or more	4.0%

Table 2: Validation Results

Condition	Chart + CPCSSN +	Chart - CPCSSN +	Chart - CPCSSN -	Chart + CPCSSN -	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
Hypertension	915	66	885	40	95.8% (94.5%, 97.1%)	93.0% (91.4%, 94.7%)	93.3% (91.7%, 94.8%)	95.7% (94.4%, 97.0%)
Diabetes	311	33	1553	9	97.2% (95.4%, 99.0%)	97.9% (97.2%, 98.6%)	90.4% (87.3%, 93.5%)	99.4% (99.0%, 99.8%)
Depression	299	47	1473	87	77.5% (73.3%, 81.6%)	96.9% (96.0%, 97.9%)	86.4% (82.8%, 90.0%)	94.4% (93.3%, 95.6%)
COPD	125	25	1730	26	82.8% (76.8%, 88.8%)	98.6% (98.0%, 99.1%)	83.3% (77.4%, 89.3%)	98.5% (98.0%, 99.1%)
Dementia	82	15	1797	12	87.2% (80.5%, 94.0%)	99.2% (98.8%, 99.6%)	84.5% (77.3%, 91.7%)	99.3% (99.0%, 99.7%)
Osteoarthritis	507	88	1211	100	83.5% (80.6%, 86.5%)	93.2% (91.9%, 94.6%)	85.2% (82.4%, 88.1%)	92.4% (90.9%, 93.8%)
Epilepsy	130	19	1751	6	95.6% (92.1%, 99.0%)	98.9% (98.4%, 99.4%)	87.3% (81.9%, 92.6%)	99.7% (99.4%, 99.9%)
Parkinsonism	66	11	1812	17	79.5% (70.8%, 88.2%)	99.4% (99.0%, 99.8%)	85.7% (77.9%, 93.5%)	99.1% (98.6%, 99.5%)

	YES	NO	UNCERTAIN	NOT APPLICABLE
TITLE, KEYWORDS, ABSTRACT				
Identify article as study of assessing diagnostic accuracy	X			
Identify article as a study of administrative data	X*			
INTRODUCTION:				
State disease identification & validation one of goals of study	X			
METHODS:				
Participants in validation cohort:				
Describe validation cohort (Cohort of patients to which reference standard was applied)				
• Age	X			
• Disease	X			
• Severity				X**
• Location/Jurisdiction	X			
Describe recruitment of validation cohort				
• Inclusion criteria	X			
• Exclusion criteria	X			
Describe patient sampling (Random, consecutive, all, etc.)				
Describe data collection				
• Who identified patients and did selection adhere to patient recruitment criteria	X			
• Who collected data	X			
• *A priori* data collection form	X			
• Disease classification	X			
• Split sample (i.e. re-validation using a separate cohort)				X
Test Methods:				
Describe number, training and expertise of persons reading reference standard	X			
If >1 person reading reference standard, quote measure of consistency (e.g. kappa)		X		

Blinding of interpreters of reference standard to results of classification by administrative data e.g. Chart abstractor blinded to how that chart was coded	X			
Statistical Methods:				
Describe methods of calculating/comparing diagnostic accuracy	X			
RESULTS:				
Participants:				
Report when study done, start/end dates of enrollment		X***		
Describe number of people who satisfied inclusion/exclusion criteria	X			
Study flow diagram		X***		
Test results:				
Report distribution of disease severity				X
Report cross-tabulation of index tests by results of reference standard	X			
Estimates:				
Report at least 4 estimates of diagnostic accuracy	X			
Diagnostic Accuracy Measures Reported:				
• Sensitivity	X			
• Spec	X			
• PPV	X			
• NPV	X			
• Likelihood ratios		X		
• Kappa		X		
• Area under the ROC curve/c-statistic		X		
• Accuracy/agreement		X		
• Other (specify)		X		
Report accuracy for subgroups (e.g. age, geography, different sex, etc.)				X
If PPV/NPV reported, ratio of cases/controls of validation	X			

cohort approximate prevalence of condition in the population				
Report 95% confidence intervals for each diagnostic measure	X			
DISCUSSION:				
Discuss the applicability of the validation findings	X			

*Our study concerns EMR-derived clinical data, rather than administrative health data

**It was not possible within this data source to adequately assess for severity of illness, so that was not a factor we included in the reference standard and have not reported distribution for severity of illness.

*** Study dates, including dates of enrollment were listed in the previous study (Williamson et al., 2014) which concerned the same sample as this study for the same period of time (patient data extracted on June 30, 2012). The study flow diagram has also been previously included in Williamson et al. (2014).