

# Manual review of electronic medical records as a reference standard for case definition development: a validation study

Tyler Williamson PhD, Rebecca C. Miyagishima MSc, Janeen D. Derochie RN, Neil Drummond PhD

## Abstract

**Background:** The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) previously carried out a validation study of case definitions for 8 chronic diseases (diabetes mellitus, hypertension, osteoarthritis, depression, dementia, chronic obstructive pulmonary disease, parkinsonism and epilepsy) using direct review of “raw” electronic medical record data. Although 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:** Using a traditional validation study design, we compared the case identification results of the chart reviews for the 8 chronic diseases with the results of a manual review of the CPCSSN processed data for the same conditions in the same patient sample. Patients were randomly sampled from the June 30, 2012 CPCSSN database, with oversampling of patients with rare conditions.

**Results:** We analyzed data for 1906 patients. Manual review of the CPCSSN records for case ascertainment yielded sensitivity ranging from 77.5% (95% confidence interval [CI] 73.3%–81.6%) for depression to 97.2% (95% CI 95.4%–99.0%) for diabetes. Specificity was high for all definitions (range 93.1% [95% CI 91.4%–94.7%] to 99.4% [95% CI 99.0%–99.8%]). Positive predictive values and negative predictive values also showed high accuracy of the manual CPCSSN record review relative to review of the raw chart data.

**Interpretation:** The use of CPCSSN records as the reference standard to validate case definitions substantially 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.

The adoption of electronic medical records (EMRs) in Canadian primary care practices provides a valuable opportunity to develop research- and surveillance-related information.<sup>1</sup> The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) is Canada’s only pan-Canadian primary care EMR database. It currently holds de-identified records for 1.7 million Canadian primary care patients from 1500 sentinel family physicians, nurse practitioners and community pediatricians in 11 provinces and territories and from 10 different EMR systems. To help ensure rigor in the surveillance of chronic disease in Canada, the CPCSSN carried out a large validation study of case definitions for 8 chronic diseases (diabetes mellitus, hypertension, osteoarthritis, depression, dementia, chronic obstructive pulmonary disease, parkinsonism and epilepsy).<sup>2</sup> The case definitions were implemented in the CPCSSN database by means of computerized algorithms that extract, clean and process the data into a standard format. The definitions were validated with the use of the accepted reference standard method of manual review of the patient’s source EMR. Vali-

ation results were favourable for all case definitions, with sensitivity ranging from 77.8% to 98.8%, specificity from 93.5% to 99.0%, positive predictive value (PPV) from 72.1% to 92.9% and negative predictive value (NPV) from 90.2% to 99.9%.

Although effective, this method of validation is time-consuming and can present challenges. Access to patient charts must be coordinated with participating clinics, entailing an increased time commitment and workload for clinic administrators. Furthermore, the potential risk to patient privacy and data security necessitates additional safeguards to be implemented by both researchers and clinic administrative

**Competing interests:** None declared.

This article has been peer reviewed.

**Correspondence to:** Tyler Williamson, tyler.williamson@ucalgary.ca

**CMAJ Open 2017. DOI:10.9778/cmajo.20170077**

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 on the effectiveness of the database in representing the information originally contained in the patients' 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 1 clinical database but may remain obscure with regard to even basic information in another. Records within the CPCSSN database have been stripped of all direct identifiers and do not contain unstructured clinical notes, referral letters or diagnostic images but otherwise contain effectively the same information as the source EMR. Therefore, we aimed to determine whether the data contained in the CPCSSN records might function as a reference standard for case definition validation.

## Methods

### Study setting

The CPCSSN is a network of 12 practice-based primary care research networks across Canada that was established in 2008.<sup>3</sup> It is Canada's first multidisease EMR surveillance system, extracting information from the EMRs of participating sentinel providers (family physicians, nurse practitioners and pediatricians) every 6 months. The CPCSSN has grown to include the de-identified EMR data for more than 1.7 million Canadians from nearly 1500 sentinel providers. Data for the current study were drawn from the June 30, 2012, CPCSSN database, collected for the primary purpose of validating the case definitions of 8 chronic conditions that have been reported elsewhere.<sup>2</sup> At that time, the CPCSSN database housed data for 600 000 patients extracted from the EMRs of 475 sentinel care providers. Our previous work has shown that CPCSSN sentinels and patients are reasonably representative of the general population of providers and patients.<sup>4</sup>

### Sample selection and design

Six of the 12 CPCSSN networks contributed to the patient sample. The 6 networks that did not participate did so for a variety of reasons: 1 was the pilot test site, 2 were not yet CPCSSN networks at the time the sample was drawn, and 3 were unable to participate because of staffing reasons.

The original sampling plan called for review of 2200 patient charts, with 5 of the networks reviewing 400 patient charts and that in British Columbia (which had a smaller number of participating sentinels) reviewing 200 charts. Of the 400 patients, 350 were randomly sampled by means of an age-stratified method, with 90% of the sample being drawn from among those more than 60 years. In addition, the sample was augmented by a random sample of 25 patients drawn from among those who were case positive for epilepsy and 25 patients drawn from among those who were case positive for parkinsonism, because the prevalence of these conditions is generally low.

### CPCSSN record review

Two experienced research assistants (R.M. and J.D., an epidemiologist and a nurse) were trained to independently review patient data records within the CPCSSN database and to assess caseness in each record separately, for each of the 8 conditions of interest. They were blinded to other assessments of caseness, including the case assignment by the CPCSSN's algorithms and the case determination made by reviewers during the original validation study. The 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. In addition, the reviewers could see both the original and cleaned text entries as well as International Classification of Diseases, 9th revision codes. Cases in which there was uncertainty were discussed with the team's lead (T.W.) 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. Consensus was reached in all such cases.

### Statistical analysis

Measures of validity used in this study were sensitivity, specificity, PPV and NPV. We calculated these by comparing the outcomes of the manual CPCSSN record review with the outcomes of the original manual review of the EMR charts (reference standard). In accordance with the methodology used by Williamson and colleagues,<sup>2</sup> we considered 70% the cut-off for validity for all measures for sensitivity and specificity. No cut-off value was assigned for PPV or NPV. We analyzed all data using Stata/IC version 13 statistical software (StataCorp).

### Ethics approval

The University of Calgary Conjoint Health Research Ethics Board approved the study.

## Results

The final sample used for our analysis included 1906 patients. The shortfall of 294 patients was due to several reasons: 149 were excluded because of EMR access issues, 87 because the EMR record was not sufficiently complete, 44 because there were not enough patients with parkinsonism in some networks to satisfy the requirement of the additional 25 patients, and 14 because the practice had left the CPCSSN before the analysis was conducted. The final sample of 1906 patients ensured a margin of error of no more than 10% for the 95% confidence intervals (CIs) for all validity dimensions.

Of the 15 248 caseness decisions (i.e., 1906 patients × 8 conditions), 347 (2.3%) were reviewed for adjudication by both reviewers and by the team lead (T.W.)

Table 1 summarizes the patients' characteristics. The sex of patients included in the study was reflective of the intention to oversample older patients, with 1063 female patients (55.8%) in the final sample. The patients' age ranged from 5

**Table 1: Patient characteristics**

Characteristic	No. (%) of patients n = 1906
Female sex	1063 (55.8)
Age > 60 yr	1630 (85.5)
Prevalence of CPCSSN chronic diseases	
Hypertension	955 (50.1)
Diabetes mellitus	320 (16.8)
Depression	386 (20.2)
Chronic obstructive pulmonary disease	151 (7.9)
Osteoarthritis	607 (31.8)
Dementia	94 (4.9)
Epilepsy	136 (7.1)
Parkinsonism	83 (4.4)
No. of CPCSSN chronic diseases	
0	428 (22.4)
1	642 (33.7)
2	515 (27.0)
3	242 (12.7)
4–8	79 (4.1)

Note: CPCSSN = Canadian Primary Care Sentinel Surveillance Network.

to 107 years, with 1630 (85.5%) being more than 60 years. Half of the patients (955 [50.1%]) had a diagnosis of hypertension noted in the chart according to the chart review. Fewer than one-quarter (428 [22.4%]) had none of the 8 conditions under study.

Table 2 outlines the results of the validation analysis. The manual review of CPCSSN records for case ascertainment yielded sensitivity ranging from 77.5% (95% CI 73.3%–81.6%) for depression to 97.2% (95% CI 95.4%–99.0%) for diabetes. Specificity was high for all definitions and ranged from 93.1% (95% CI 91.4%–94.7%) for hypertension to 99.4% (95% CI 99.0%–99.8%) for parkinsonism. The PPV and NPV showed the manual record review to be highly accurate: the PPV ranged from 83.3% (95% CI 77.4%–89.3%) for chronic obstructive pulmonary disease to 93.3% (95% CI 91.7%–94.8%) for hypertension, and the NPV ranged from 92.4% (95% CI 90.9%–93.8%) for osteoarthritis to 99.7% (95% CI 99.4%–99.9%) for epilepsy. Overall, the case definition for diabetes achieved the highest sensitivity and specificity (97.2% [95% CI 95.4%–99.0%] and 97.9% [95% CI 97.2%–98.6%], respectively).

### Interpretation

The results of this study show 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), whereas con-

**Table 2: Results of validation analysis using the outcome of original manual review of the electronic medical record charts as the reference standard**

Disease	No. of cases				Sensitivity (95% CI), %	Specificity (95% CI), %	Positive predictive value (95% CI), %	Negative predictive value (95% CI), %
	Chart +, CPCSSN +	Chart –, CPCSSN +	Chart –, CPCSSN –	Chart +, CPCSSN –				
Hypertension	915	66	885	40	95.8 (94.5–97.1)	93.1 (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)
Chronic obstructive pulmonary disease	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)	87.2 (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)

Note: CI = confidence interval, CPCSSN = Canadian Primary Care Sentinel Surveillance Network.

ditions 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 and colleagues<sup>5</sup> examined billing code and free-text diagnoses of upper gastrointestinal bleeding in 2 primary care databases (as well as 2 administrative databases) based in the Netherlands (Integrated Primary Care Information) and Italy (Health Search/CSD Patient Database). Positive predictive values ranged from 21% for the former to 78% for the latter. John and colleagues<sup>6</sup> validated Read codes for anxiety and depression in a Welsh primary care health record database linked with survey results from a community health inequality survey. They reported insufficient validity, with sensitivity ranging from 0.05% to 0.49%. The EMERALD database in Ontario, based on a single primary care EMR system, has been used as a reference standard for several studies validating case definitions;<sup>7-10</sup> 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 that, by convention, is generally considered to be preeminent: the medical chart. We are aware of no previous work seeking to rigorously validate case definitions developed with the use of processed EMR data in comparison to those formulated with the use of the conventional reference standard, particularly when linkage to nonprimary care data is not feasible. This study therefore represents a substantial contribution to both primary care and health information technology research in the Canadian context. This has major significance for the development of future CPCSSN case definitions, as it will allow researchers to streamline the work and will dramatically reduce time and cost constraints, which previously presented challenges. This work should lead to substantial increases in the number of conditions in the CPCSSN data with validated case definitions, yielding improvement in the utility of the data for research, surveillance and quality-improvement studies.

### Limitations

There are limitations to using an EMR database as a reference standard. Data derived from EMRs are 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 using the data.<sup>3</sup> Several records in our study had to be excluded owing 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 from the charts by the CPCSSN. Unstructured clinical notes, referral letters and diagnostic images are among those not extracted from the source EMR record for reasons of confidentiality. If deterministic information is expected to be found in that type of data,

a manual review of the CPCSSN data will not serve well as a reference standard. The lower validity values for osteoarthritis and depression we report here may be examples of this effect, although the values are typically judged as being acceptable.

### Conclusion

Our results show that chart reviews, which are often challenging owing to time and financial constraints, are a sufficient but sometimes unnecessary reference standard. The use of CPCSSN record data to validate case definitions substantially reduces the burden on sentinel physicians and clinic managers as well as researchers. This shorter, more cost-effective process for case definition validation should increase the potential for future work validating case definitions 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* 2015;187:239-40.
2. Williamson T, Green ME, Birtwhistle R, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med* 2014;12:367-72.
3. Birtwhistle R, Keshavjee K, Lambert-Lanning A, et al. Building a pan-Canadian primary care sentinel surveillance network: initial development and moving forward. *J Am Board Fam Med* 2009;22:412-22.
4. Queenan JA, Williamson T, Khan S, et al. Representativeness of patients and providers in the Canadian Primary Care Sentinel Surveillance Network: a cross-sectional study. *CMAJ Open* 2016;4:E28-32.
5. Valkhoff VE, Coloma PM, Masclee GMC, et al.; 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. *J Clin Epidemiol* 2014;67:921-31.
6. John A, McGregor J, Fone D, et al. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;16:35.
7. Butt DA, Tu K, Young J, et al. A validation study of administrative data algorithms to identify patients with parkinsonism with prevalence and incidence trends. *Neuroepidemiology* 2014;43:28-37.
8. Widdifield J, Bombardier C, Bernatsky S, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord* 2014;15:216.
9. Schwartz KL, Tu K, Wing L, et al. Validation of infant immunization billing codes in administrative data. *Hum Vaccin Immunother* 2015;11:1840-7.
10. Tu K, Wang M, Young J, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMERALD as a reference standard. *Can J Cardiol* 2013;29:1388-94.

**Affiliations:** Community Health Sciences (Williamson, Derochie), University of Calgary, Calgary, Alta.; School of Public Health (Miyagishima), University of Alberta, Edmonton, Alta.; Department of Family Medicine (Drummond), University of Alberta, Edmonton, Alta.; Department of Family Medicine (Drummond), University of Calgary, Calgary, Alta.

**Contributors:** Tyler Williamson and Neil Drummond conceived of and designed the project and drafted the manuscript. Rebecca Miyagishima and Janeen Derochie contributed to the development of the manual review methodology, acquired the data, contributed to manuscript writing and reviewed the manuscript for important intellectual content. Tyler Williamson analyzed the data. All of the authors contributed to interpretation of the data, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

**Supplemental information:** For reviewer comments and the original submission of this manuscript, please see [www.cmajopen.ca/content/5/4/E830/suppl/DC1](http://www.cmajopen.ca/content/5/4/E830/suppl/DC1).