Quality of administrative health databases in Canada: A scoping review

Aynslie Hinds, MSc,¹ Lisa M. Lix, PhD,¹ Mark Smith, MSc,² Hude Quan, PhD,³ Claudia Sanmartin, PhD⁴

ABSTRACT

OBJECTIVE: Administrative health databases are increasingly used to conduct population-based health research and surveillance; this has resulted in a corresponding growth in studies about their quality. Our objective was to describe the characteristics of published Canadian studies about administrative health database quality.

METHODS: PubMed, Scopus, and Google Advanced were searched, along with websites of relevant organizations. English-language studies that evaluated the quality of one or more Canadian administrative health databases between 2004 and 2014 were selected for inclusion. Extracted information included data quality concepts and measures, year and type of publication, type of database, and geographic origin.

SYNTHESIS: More than 3,000 publications were identified from the search. Twelve reports and 144 peer-reviewed papers were included. The majority (53.5%) of peer-review publications used databases from Ontario and Alberta, while 67% of the non-peer-review publications used data from multiple provinces/territories. Almost all peer-reviewed papers (97.2%) were validation studies. Hospital discharge abstracts and physician billing claims were the most frequently validated databases. Approximately half of the publications (53.0%) validated case definitions and 37.7% focused on a chronic physical health condition.

CONCLUSION: Gaps in the Canadian administrative data quality literature include a limited number of studies evaluating data from the Maritimes and across multiple jurisdictions, newer data sources, validating methods for identifying individuals with mental illness, and assessing the completeness and serviceability of the data. Data quality studies can aid researchers to understand the strengths and limitations of the data.

KEY WORDS: Data linkage; administrative health database; diagnosis codes; validation studies; review study

La traduction du résumé se trouve à la fin de l'article.

Can J Public Health 2016;107(1):e56–e61
doi: 10.17269/CJPH.107.5244

Canada has one of the richest collections worldwide of electronic health databases, including administrative health databases, clinical registries, electronic medical records, and health surveys.^{1,2} The number of studies that use electronic health databases for research and surveillance about population health, health service use, and the determinants of health, is rapidly increasing, as are the variety of available databases and number of cross-provincial investigations. In this study, we focus specifically on administrative health databases, which include hospital records, physician billing claims, population registries, and prescription drug records. Administrative health databases have long been used for chronic and infectious disease research and surveillance and are the foundation for such initiatives as the Canadian Chronic Disease Surveillance System (CCDSS) and the Canadian Network of Observational Drug Effect Studies (CNODES).^{3,4}

Administrative health databases were originally created for health care management and monitoring functions, such as remunerating physicians. Although these databases were not originally designed to be used for research and surveillance, they contain a rich source of information that is now routinely used for these purposes.^{5,6} Therefore a central question underlying studies that use administrative health databases is “Are the data of good quality for their intended use?” Given that administrative data are often used in epidemiological and public health research, it is

important to conduct data quality studies so researchers are kept informed about the strengths and limitations of the data, and can then take steps to produce unbiased results by minimizing selection bias and measurement error. These efforts support good decision making about health and health care use.

Data quality is a broad concept that is both relative and multidimensional.⁷ One comprehensive definition is “the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data”.⁸ The multidimensional nature of data quality is evident in conceptual frameworks developed by national/provincial agencies and organizations that either generate or use electronic health databases for research and surveillance. For example, Canadian Institute for Health Information’s (CIHI) data quality framework

Author Affiliations

1. Department of Community Health Sciences, University of Manitoba, Winnipeg, MB
2. Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, MB
3. Department of Community Health Sciences, University of Calgary, Calgary, AB
4. StatsCan, Health Analysis Division, Ottawa, ON

Correspondence: Lisa Lix, PhD, Department of Community Health Sciences, University of Manitoba, 5113-750 Bannatyne Avenue, Winnipeg, MB R3E 0W3, Tel: 204-789-3573, E-mail: Lisa.Lix@umanitoba.ca

Acknowledgement: The authors thank Shelley-May Neufeld for her contributions to conducting the literature search and developing the data extraction form.

Funding: LML is currently supported by a Manitoba Health Research Chair; she was supported by a University of Saskatchewan Centennial Research Chair at the time this research was initiated. HQ is supported by Alberta Innovate Health Solution.

Conflict of Interest: None to declare.

encompasses concepts of accuracy, timeliness, comparability, usability and relevance.⁹ The Public Health Agency of Canada's (PHAC) framework encompasses similar concepts, including accuracy, timeliness, serviceability, usability and relevance. The Statistics Canada framework includes the concepts of relevance, accuracy, timeliness, accessibility, interpretability and coherence.¹⁰ The latter concept refers to the ability to bring together data from different sources; for example, coherence can be achieved by using common methods across surveys or standardized coding systems across time or geography. The Manitoba Centre for Health Policy (MCHP) and the Institute for Clinical Evaluative Services (ICES) have also produced data quality frameworks.^{5,11} The measurement of these data quality concepts is addressed in the frameworks.

Currently, there is no summary of recent Canadian administrative health data quality studies that could inform future research in this field. Our objective was to review existing research that has described the quality of Canadian administrative health databases, and to examine the characteristics of data quality studies using these frameworks as a guide.¹² Our goal was to identify gaps in existing research and pinpoint new opportunities to evaluate and improve our understanding of administrative health database quality.

METHODS

Data source

A scoping review was conducted; it was based on a comprehensive literature search of published peer-reviewed and non-peer-reviewed Canadian studies. We searched two electronic citation indices – PubMed and Scopus – using two groups of search terms about administrative data and data quality concepts (see Table 1) linked with Boolean operators. Canadian studies were identified by specifying “Canada” as the affiliation country of the author(s).

Table 1. Scoping review search parameters: Search terms and relevant organizational websites

Search terms for administrative data	Search terms for data quality
administrative data health administrative data administrative health data administrative health database registry hospital discharge abstracts physician claims	validity OR valid accuracy OR accurate correctness OR correct reliability OR reliable data quality comprehensiveness OR comprehensive anonymity linkability usability timeliness OR temporal consistency
Name of organization	Website
Canadian Institute for Health Information	http://www.cihi.ca
Centre for Health Services and Policy Research	http://www.chspr.ubc.ca/
Health Council of Canada	http://www.healthcouncilcanada.ca/
Health Quality Council of Alberta	http://www.hqca.ca/
Institute for Clinical Evaluative Sciences	http://www.ices.on.ca/
Manitoba Centre for Health Policy	http://www.umanitoba.ca/centres/mchp
Newfoundland & Labrador Centre for Health Information	http://www.nlchi.nl.ca/
Population Health Research Unit	http://www.phru.dal.ca/index.cfm
Saskatchewan Health Quality Council	http://www.hqc.sk.ca
Statistics Canada	http://www.statcan.gc.ca

We also searched for relevant publications using Google Advanced, in the reference lists of the included articles, and on websites of relevant national and provincial organizations and agencies (Table 1). The review was limited to studies published between 2004 and 2014 because a review of data quality studies conducted prior to 2004 has already been published.¹³

Selection and data abstraction

One investigator (AH) reviewed all of the publications identified via the literature search. The second author (LL) reviewed a 30% sample of the articles to ensure that the inclusion and exclusion criteria were consistently applied and that the information abstracted was correctly classified. A publication was excluded if the data source that was evaluated was not an administrative health database or if it was not classified as a study about data quality. In this review, administrative health data sources included health insurance registries, hospital discharge abstracts, physician billing claims, emergency department records, electronic medical records, and drug and vaccine information systems.¹⁴ Disease-specific registries were also included. Only electronic data sources were included in this review.

Data abstracted from each publication retained in the scoping review were systematically recorded in a database. Data included publication year and type of publication (i.e., peer-reviewed, non-peer reviewed). The geographic origin of the database(s) and type of database(s) evaluated in each publication were noted.

Publications were classified as validation studies or “other” studies. Validation studies examine the accuracy of diagnoses, procedures, or other health measures by comparing them to another, unbiased data source.¹⁵ Validation studies were categorized as ecological, re-abstraction, or reference standard using van Walraven and Austin's (2012) criteria.¹⁶ Ecological studies compare statistics about event rates calculated from administrative data with rates obtained from a valid data source. Re-abstraction studies involve a re-collection of data from the original source (i.e., patient chart); the re-abstracted data are compared with the administrative data to assess validity.¹⁷ Reference standard validation studies compare study data to an error-free reference standard data source, such as clinical criteria, a panel consensus review, or medical charts.¹⁵ Studies categorized as “other” included all data quality studies that were not classified as validation studies, such as those that examined the feasibility of linking databases.

The data quality concept(s) that were the focus of the study were noted, specifically correctness/reliability, completeness and serviceability. These data quality concepts were selected from the CIHI, PHAC, MCHP, Statistics Canada and ICES data quality frameworks. All of the data quality frameworks include accuracy, a multicomponent concept which involves correctness/reliability and completeness. Serviceability is an element of PHAC's data quality framework, which includes linkability and temporal consistency found in other data quality frameworks.

For validation studies, the element of the data for which validity was evaluated (i.e., diagnosis codes, intervention or procedure codes, algorithm or case definition, other measure) was noted. The health condition(s) of interest in the quality evaluation were classified as: chronic physical health condition (e.g., diabetes, chronic obstructive pulmonary disease, cancer), acute physical health condition (e.g., acute coronary syndrome, acute myocardial infarction), mental health condition (e.g., depression,

schizophrenia), multiple conditions, or not applicable (e.g., screening test, surgery, vaccination). The data quality measure(s) (i.e., sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) were recorded. Descriptive statistics, including frequencies and percentages, were used to analyze the data.

SYNTHESIS

More than 3,000 peer-review publications and non-peer-review publications were identified via searches of the citation indices and organizational websites. There was a high degree of consistency between the two authors (AH and LL) in applying the inclusion/exclusion criteria and in categorizing the abstracted information.

Figure 1 is a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart that shows the number of articles included and excluded at each stage of the review process. Articles were excluded for various reasons: the data source under review was not an electronic administrative database (e.g., medical charts),^{18–20} the data were not from Canada,^{21,22} or the article was a summary of other literature.^{23,24} Four articles were excluded because the abstract and article were not published in English (two in French, one in Spanish and one in German). A total of 156 publications were retained, of which 144 (92.3%) were peer-reviewed publications and 12 (7.7%) were non-peer-reviewed publications (see Supplementary Table 1 in the ARTICLE TOOLS section on the journal site).

Year of publication and geographic origin of the database(s)

As Figure 2 reveals, the frequency of Canadian publications about the quality of administrative health databases increased over the study period, from 4 publications in 2004 to 20 publications in 2014. The majority of peer-reviewed publications were based on

Figure 1. PRISMA flow chart of articles included and excluded in this scoping review

Figure 2. Number of Canadian data quality publications by year and region of the database

administrative health databases from Ontario (34.0%) and Alberta (19.4%). The majority of non-peer-reviewed publications were based on databases from multiple provinces or territories (66.7%). For example, 6 of the 12 non-peer-reviewed publications were CIHI studies investigating the Hospital Discharge Abstract Database across provinces and territories.

Validation studies

Type of Study, Study Design, and Data Quality Concepts

Overall, 96.8% of the publications were classified as validation studies. The majority of these validation studies used a reference standard study design (83.4%), although ecological (15.9%), and re-abstraction (6.0%) studies were also identified. The majority (75.0%) of the non-peer-reviewed publications were re-abstraction studies. Among the validation studies, 95.4%, 9.3% and 6.0% addressed the correct/reliable, complete, and serviceable data quality concepts respectively.

Databases Validated

Hospitalization (70.2%), physician billing claims (60.3%) and emergency department records (17.9%; Table 2) were the most commonly validated databases. Frequently, multiple data sources were validated simultaneously (i.e., 46.4% validated one data source, 25.8% validated two data sources, and 27.8% validated three or more data sources). When multiple databases were validated, one was usually the hospitalization database. Medical charts were the most common reference standard data source (53.2%). However, clinical registries, self-report data from the National Population Health Survey and Canadian Community Health Survey, and electronic medical records were also used as the reference standard.²⁵

Subject of Validation

More than one half of the validation studies conducted an investigation of a disease algorithm or case definition (53.0%),

Table 2. Characteristics of Canadian validation studies (*n* = 151) of electronic health databases

Data source	<i>n</i> (%)
Hospital discharge abstracts	106 (70.2)
Physician billing claims	91 (60.3)
Emergency department records	27 (17.9)
Prescription drug records	27 (17.9)
Surveillance system or registry	25 (16.6)
Mental health information system	4 (2.6)
Vital statistics	1 (0.7)
Electronic medical records	8 (5.3)
Minimum data set	3 (2.0)
Topic of validation	
Diagnosis codes	35 (23.2)
Intervention or procedure codes	16 (10.6)
Disease algorithm or case definition	80 (53.0)
Prescription codes	2 (1.3)
Other measure (e.g., comorbidity index)	37 (24.5)
Health condition	
Chronic physical health	57 (37.7)
Acute physical health	17 (11.3)
Mental health	6 (4.0)
Multiple conditions	38 (25.2)
Other*	48 (31.8)
Measure of validity	
Sensitivity	97 (64.2)
Specificity	83 (55.0)
Negative predictive value	65 (43.0)
Positive predictive value	88 (58.3)
Kappa	46 (30.5)
Likelihood ratio statistic	8 (5.3)
Other (e.g., c-statistic)	84 (55.6)

* Includes demographic characteristics and measures of health service use, such as length of hospital stay, number of hospitalizations, receipt of care from a specialist, surgical wait times, vaccination status, receipt of screening tests, receipt of a transplant, type of treatment/procedure/intervention, and location of treatment.

while almost one quarter (23.2%) investigated diagnosis codes and 10.6% investigated intervention or procedure codes. Case definitions were often based on diagnoses identified in multiple data sources (e.g., hospital discharge abstracts and physician billing claims; see Table 2).

Health Condition Validated

Chronic physical conditions were most commonly validated, including asthma, cancer, diabetes, hypertension, ischemic heart disease, osteoporosis, chronic obstructive pulmonary disease, epilepsy, irritable bowel syndrome, and the human immunodeficiency virus. Often multiple disease case definitions were examined; for example, Marrie et al.'s study validated case definitions for hypertension, hyperlipidemia and diabetes among individuals with multiple sclerosis.²⁶ As another example, Cadieux and Tamblyn focused on several different types of acute respiratory infections.²⁷ Other elements of administrative data were validated; Monfared et al. used medical claims to estimate hospital length of stay (using hospital data as the reference standard) and DeCoster et al. examined the accuracy of surgical wait times.^{28,29}

Data Quality Measures

Sensitivity and PPV were the most common measures of validity, used by 64.2% and 58.3% of validation studies respectively. Specificity and NPV were also used frequently. Kappa was used when neither data source was considered the reference standard. A follow-up analysis of the characteristics of these validation studies

by the geographic region of the study databases was conducted (see Supplementary Table 2).

"Other" studies

Only 3.2% of the studies were not classified as validation studies. Among these studies, different data quality topics were investigated (i.e., completeness and serviceability) and all used basic descriptive statistics (i.e., percentages, averages) to analyze the data. For example, Cunningham et al. examined the face validity of the physician billing claims database by examining the average number of claims per physician within a three-month period, and the average number of unique diagnosis codes by physician specialty and by payment method (i.e., fee-for-service, salaried).³⁰ Alshammari and Hux focused on the completeness of diabetes prevalence estimates from administrative databases, considering the effect of alternate methods of physician payment on the use of physician billing claims databases to ascertain diabetes cases.³¹ Metcalfe et al. examined the feasibility of linking several databases to identify a cohort of pregnant women based on the date of conception (rather than the date of birth) in order to capture fetal loss.³² Li et al. examined the accuracy of linking three databases (vital statistics, population registry, hospital discharge abstracts) using different combinations of common identifiers.³³

DISCUSSION

We found the number of data quality publications on administrative health databases in Canada increased over time, with 156 publications in an 11-year period being included in this scoping review. Hospitalization and physician billing claims data were the most common validated databases; however, the diversity of data sources that are being evaluated is increasing. Most studies using national-level data were undertaken by CIHI. CIHI employs a set of quality control measures to ensure high-quality hospitalization data, including applying 900 data element edits to each abstract and examining relationships between data elements.³⁴

Most of the studies included in the scoping review were classified as validation studies and examined correctness. The majority used a reference standard approach, commonly using medical charts as the reference standard, though clinical registries and surveys were also used. Disease algorithms/case definitions, often for chronic conditions, were the most common subjects of investigation, which is consistent with findings from Roos et al.'s literature review.¹³ Diagnostic codes and other measures, such as comorbidity indices, were also frequently examined. Sensitivity and PPV were the most common validity measures, which is also consistent with other reviews of administrative health data quality.^{35,36}

There are a number of gaps in the current body of Canadian studies validating administrative health data. First, the majority of the studies examined databases from Alberta and Ontario. There have been no data quality studies focused solely on provincial/territorial-level data from New Brunswick, Prince Edward Island, the Northwest Territories and Nunavut; however, these jurisdictions are included in some national studies. Also, there is a lack of peer-reviewed studies involving data from multiple provinces/territories. Thus, there may be disparity in data quality knowledge across Canada which could have a negative impact on cross-jurisdictional initiatives. More than a decade ago, Kephart noted the difficulty in accessing and analyzing administrative

health databases in multiple provinces.³⁷ Efforts to improve access should increase the number of multi-jurisdictional studies.

Second, there were very few publications that reported on the quality of electronic medical records, the Resident Assessment Instrument Minimum Dataset (RAI-MDS),^{38,39} or population registries. Electronic medical records are used in primary care settings to record and electronically store patient health information, including medical histories, laboratory results and prescriptions, and are increasingly being used for research and surveillance.^{40,41} The RAI-MDS is a standardized tool used to assess residents of long-term care facilities and the information collected is used to create quality-of-care indicators.⁴² Population registries capture demographic and residence location information about health insurance beneficiaries and are of central importance to all population-based investigations. Roos et al. recommended that data quality studies about population registries focus on identifiers of client residence location, including those that can be used to track client location over time; the quality of residence location information is important for studies about area-based socio-economic factors and geographic determinants of health and health care use.¹³

Third, for each study we determined which data quality concept(s), from the CIHI, PHAC, Statistics Canada, MCHP and ICES data quality frameworks, were applied. Most publications investigated correctness (diagnostic accuracy, bias, measurement error); few publications examined other facets of data quality, including the completeness of a database (coverage, amount of missing information) and serviceability (i.e., linkability of data sources, consistency over time). Thus, there is a need for more data quality studies examining these data quality concepts to enhance the growing body of knowledge in this field. For example, we support Roos et al.'s recommendation for more studies about the completeness and comparability of administrative data for capturing prescription dispensations.¹³ Such studies would be useful because changes in provincial formularies, the list of prescription drugs covered by health insurance benefits, may affect the usefulness of the data for longitudinal studies of drug exposures.

There is no consistently used reference standard data source in validation studies that used a reference standard design; choosing an appropriate one is challenging and may vary depending on the purpose of the study.⁴³ Variations in findings across validation studies may be partly explained by differences in the reference standard.^{36,44}

Last, there is a dearth of studies that have validated methods for identifying individuals with mental health and acute health conditions using administrative health data. In some jurisdictions, multiple coding systems are used (i.e., ICD-9-CM, ICD-10-CA). Few studies have assessed the accuracy of diagnosis codes or case definitions across coding systems.³⁶

Limitations and strengths

This scoping review has some limitations. We only reviewed articles written in English; however, there were only four studies published in a different language. Moreover, 15.2% of the validation studies included in our scoping review were published in English and validated administrative health databases from Quebec, a predominantly French-speaking province. We did not compare the data quality studies conducted on Canadian data with

those conducted on administrative health databases from other countries, information that might have helped to provide some context to identify gaps in the data quality research literature. The scoping review did not encompass all types of electronic health databases; we focused on administrative health databases because they are so widely used. The strengths of this study are the comprehensive search of both peer-reviewed and non-peer-reviewed publications, and the wide range of information that was collected from each publication.

A future scoping review could extend the current one by including administrative health data quality studies from other countries to do a cross-country comparison. Additionally, a future review could evaluate the extent to which administrative health data quality studies evaluate other data quality concepts included in the various data quality frameworks, such as usability and timeliness.

CONCLUSIONS

This scoping review identified numerous gaps in the Canadian administrative data quality literature and identified opportunities to enhance or improve it. Administrative data have multiple uses – patient care, administrative functions, surveillance, and conducting of policy-relevant research, upon which decisions about population health and health services are made.⁴⁵ Efforts to address these gaps will increase our understanding of the strengths and limitations of these data for all their uses.

REFERENCES

1. Bailie L, Dufour J, Hamel M. *Data Quality Assurance for the Canadian Community Health Survey (CCHS)*. Ottawa, ON: Statistics Canada, 2002.
2. Das B, Clegg LX, Feuer EJ, Pickle LW. A new method to evaluate the completeness of case ascertainment by a cancer registry. *Cancer Causes Control* 2008;19:515–25. PMID: 18270798. doi: 10.1007/s10552-008-9114-0.
3. Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L. Mining administrative health databases to advance medical science: Geographical considerations and untapped potential in Canada. *Can J Cardiol* 2012;28:152–54. PMID: 22301469. doi: 10.1016/j.cjca.2012.01.005.
4. Shah BR, Lipscombe LL. Clinical diabetes research using data mining: A Canadian perspective. *Can J Diabetes* 2015;39:235–38. PMID: 26004906. doi: 10.1016/j.cjcd.2015.02.005.
5. Iron K, Manuel DG. Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data. Toronto, ON, 2007. Available at: <http://www.ices.on.ca/~media/Files/Atlases-Reports/2007/Quality-assessment-of-administrative-data/Full%20report.ashx> (Accessed May 30, 2016).
6. Wray NP, Ashton CM, Kuykendall DH, Hollingsworth JC. Using administrative databases to evaluate the quality of medical care: A conceptual framework. *Soc Sci Med* 1995;40:1707–15. doi: 10.1016/0277-9536(94)00275-X.
7. Statistics Canada. Statistics Canada Quality Guidelines. Ottawa, ON, 2009. Available at: <http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf> (Accessed July 8, 2015).
8. Arts DGT, De Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: A literature review, case study, and generic framework. *J Am Med Inform Assoc* 2002;9:600–11. doi: 10.1197/jamia.M1087.
9. Canadian Institute for Health Information. The CIHI Data Quality Framework. Ottawa, ON, 2009. Available at: https://www.cihi.ca/en/data_quality_framework_2009_en.pdf (Accessed July 15, 2015).
10. Statistics Canada. Statistics Canada's Quality Assurance Framework. Ottawa, ON, 2002. Available at: <http://www.statcan.gc.ca/pub/12-586-x/12-586-x2002001-eng.pdf> (Accessed May 15, 2015).
11. Azimae M, Smith M, Lix L, Ostapuk T, Burchill C, Hong SP. MCHP Data Quality Framework. Winnipeg, MB: Manitoba Centre for Health Policy, 2013. Available at: http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departamental_units/mchp/protocol/media/Data_Quality_Framework.pdf (Accessed March 15, 2015).
12. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19–32. doi: 10.1186/1748-5908-5-69.
13. Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging* 2005; 24:153–70. doi: 10.1353/cja.2005.0055.

14. Spasoff R. *Epidemiologic Methods for Health Policy*. New York, NY: Oxford University Press, 1999.
15. Spiegelman D. Validation study. In: Armitage P, Colton T (Eds.), *Encyclopedia of Biostatistics*. West Sussex: Wiley, 2005; pp. 5656–63. doi: 10.1002/0470011815.b2a03128.
16. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol* 2012;65:126–31. PMID: 22075111. doi: 10.1016/j.jclinepi.2011.08.002.
17. Richards H. Recent data quality initiatives at CIHI. In: *Ontario Health Information Management Association (OHIMA)*, 2006. Available at: http://www.powershow.com/view1/253b67-ZDc1Z/Recent_Data_Quality_Initiatives_at_CIHI_powerpoint_ppt_presentation (Accessed July 20, 2015).
18. Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. *Thromb Res* 2006;118:253–62. PMID: 16081144. doi: 10.1016/j.thromres.2005.06.015.
19. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424–41. PMID: 24843434. doi: 10.1111/j.1475-6773.2007.00822.x.
20. Blackburn DF, Schnell G, Lamb DA, Tsuyuki RT, Stang MR, Wilson TW. Coding of heart failure diagnoses in Saskatchewan: A validation study of hospital discharge abstracts. *J Popul Ther Clin Pharmacol* 2011;18:e407–15. PMID: 21900705.
21. Sears JM, Bowman SM, Hogg-Johnson S, Shorter ZA. Linkage and concordance of trauma registry and hospital discharge records. *J Occup Environ Med* 2014; 56:878–85. PMID: 25099416. doi: 10.1097/JOM.0000000000000198.
22. Waikar SS, Wald R, Chertow GM, Curhan GC, Winkelmayer WC, Liangos O, et al. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification Codes for Acute Renal Failure. *J Am Soc Nephrol* 2006; 17:1688–94. PMID: 16641149. doi: 10.1681/ASN.2006010073.
23. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health* 2011;32:91–108. PMID: 21219160. doi: 10.1146/annurev-publhealth-031210-100700.
24. Quach S, Blais C, Quan H. Administrative data have high variation in validity for recording heart failure. *Can J Cardiol* 2010;26:e306–12. doi: 10.1016/S0828-282X(10)70438-4.
25. Lix L, Yogendran M, Burchill C, Metge C, Mckeen N, Moore D, et al. *Defining and Validating Chronic Diseases: An Administrative Data Approach*. Winnipeg, MB: Manitoba Centre for Health Policy, 2006.
26. Marrie RA, Yu BN, Leung S, Elliott L, Caetano P, Warren S, et al. Rising prevalence of vascular comorbidities in multiple sclerosis: Validation of administrative definitions for diabetes, hypertension, and hyperlipidemia. *Mult Scler J* 2012;18:1310–19. PMID: 22328682. doi: 10.1177/1352458512437814.
27. Cadieux G, Tamblyn R. Accuracy of physician billing claims for identifying acute respiratory infections in primary care. *Health Serv Res* 2008;43:2223–38. PMID: 21211054. doi: 10.1186/1471-2458-11-17.
28. Monfared AAT, Leloir J. Accuracy and validity of using medical claims data to identify episodes of hospitalizations in patients with COPD. *Pharmacoeconom Drug Saf* 2006;15:19–29. doi: 10.1002/(ISSN)1099-1557.
29. De Coster C, Luis A, Taylor MC. Do administrative databases accurately measure waiting times for medical care? Evidence from general surgery. *Can J Surgery* 2007;50:394–96. PMID: 18031641.
30. Cunningham CT, Cai P, Topps D, Svenson LW, Jetté N, Quan H. Mining rich health data from Canadian physician claims: Features and face validity. *BMC Res Notes* 2014;7:682. PMID: 25270407. doi: 10.1186/1756-0500-7-682.
31. Alshammari AM, Hux J. The impact of non-fee-for-service reimbursement on chronic disease surveillance using administrative data. *Can J Public Health* 2009;100:472–74. PMID: 20209744.
32. Metcalfe A, Lyon AW, Johnson J-A, Bernier F, Currie G, Lix LM, et al. Improving completeness of ascertainment and quality of information for pregnancies through linkage of administrative and clinical data records. *Ann Epidemiol* 2013;23:444–47. PMID: 23790349. doi: 10.1016/j.annepidem.2013.05.002.
33. Li B, Quan H, Fong A, Lu M. Assessing record linkage between health care and vital statistics databases using deterministic methods. *BMC Health Serv Res* 2006;6:48. PMID: 16597337. doi: 10.1186/1472-6963-6-48.
34. Canadian Institute for Health Information. Data Quality Documentation for External Users: Discharge Abstract Database, 2010–2011. Ottawa, ON, 2011. Available at: https://www.cihi.ca/en/dad_executive_sum_10_11_en.pdf (Accessed July 20, 2015).
35. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–29. PMID: 21194889. doi: 10.1016/j.jclinepi.2010.10.006.
36. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: A systematic review. *PLoS One* 2014;9:e92286. PMID: 24682186. doi: 10.1371/journal.pone.0092286.
37. Kephart G. Barriers to Accessing and Analyzing Health Information in Canada. Ottawa, ON, 2002. Available at: https://secure.cihi.ca/free_products/CPHI_Barriers_e.pdf (Accessed July 15, 2015).
38. Foebe AD, Hirdes JP, Heckman GA, Kergoat M-J, Patten S, Marrie RA. Diagnostic data for neurological conditions in interRAI assessments in home care, nursing home and mental health care settings: A validity study. *BMC Health Serv Res* 2013;13:457. PMID: 24176093. doi: 10.1186/1472-6963-13-457.
39. Lix LM, Yan L, Blackburn D, Hu N, Schneider-Lindner V, Teare GF. Validity of the RAI-MDS for ascertaining diabetes and comorbid conditions in long-term care facility residents. *BMC Health Serv Res* 2014;14:17. PMID: 24423071. doi: 10.1186/1472-6963-14-17.
40. Birtwhistle R, Williamson T. Primary care electronic medical records: A new data source for research in Canada. *Can Med Assoc J* 2015;187:239–40. PMID: 25421989. doi: 10.1503/cmaj.140473.
41. Wasserman RC. Electronic medical records (EMRs), epidemiology, and epistemology: Reflections on EMRs and future pediatric clinical research. *Acad Pediatr* 2011;11:280–87. PMID: 21622040. doi: 10.1016/j.acap.2011.02.007.
42. Hutchinson AM, Milke DL, Maisey S, Johnson C, Squires JE, Teare G, et al. The Resident Assessment Instrument-Minimum Data Set 2.0 quality indicators: A systematic review. *BMC Health Serv Res* 2010;10:166. PMID: 20550719. doi: 10.1186/1472-6963-10-166.
43. Hudson M, Avina-Zubieta A, Lacaille D, Bernatsky S, Lix L, Jean S. The validity of administrative data to identify hip fractures is high – A systematic review. *J Clin Epidemiol* 2013;66:278–85. PMID: 23347851. doi: 10.1016/j.jclinepi.2012.10.004.
44. Jolley RJ, Sawka KJ, Yergens DW, Quan H, Jetté N, Doig CJ. Validity of administrative data in recording sepsis: A systematic review. *Crit Care* 2015; 19:139. PMID: 25887596. doi: 10.1186/s13054-015-0847-3.
45. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004; 42:1066–72. doi: 10.1097/00005650-200411000-00005.

Received: July 30, 2015
Accepted: November 19, 2015

RÉSUMÉ

OBJECTIF : On utilise de plus en plus les bases de données administratives sur la santé dans la recherche et la surveillance populationnelles en santé; le nombre d'études sur la qualité de ces bases de données croît lui aussi. Nous avons cherché à décrire les caractéristiques des études canadiennes publiées portant sur la qualité des bases de données administratives sur la santé.

MÉTHODE : Nous avons interrogé PubMed, Scopus et Google Advanced, ainsi que les sites Web d'organismes pertinents. Nous avons inclus les études en anglais évaluant la qualité d'une ou de plusieurs bases de données administratives canadiennes sur la santé entre 2004 et 2014. Nous en avons extrait : les concepts et les indicateurs de la qualité des données; l'année et le type de publication; le type de base de données; et l'origine géographique.

SYNTHÈSE : La recherche a permis de répertorier plus de 3 000 publications. Douze rapports et 144 communications évaluées par des pairs ont été inclus. La majorité (53,5 %) des publications à comité de lecture utilisaient des bases de données de l'Ontario et de l'Alberta, tandis que 67 % des publications sans comité de lecture utilisaient des données de plusieurs provinces ou territoires. Presque toutes les communications évaluées par des pairs (97,2 %) étaient des études de validation. Les registres des sorties des hôpitaux et les demandes de paiement des médecins étaient les bases de données les plus fréquemment validées. Environ la moitié des publications (53,0 %) validaient des définitions de cas et 37,7 % portaient sur un problème de santé physique chronique.

CONCLUSION : La documentation sur la qualité des données administratives canadiennes comporte des lacunes, car un nombre limité d'études évaluent les données des provinces maritimes ou de plusieurs provinces ou territoires; évaluent les nouvelles sources de données; valident les méthodes d'identification des personnes atteintes de maladies mentales; et évaluent l'exhaustivité et la fonctionnalité des données. Les études sur la qualité des données peuvent aider les chercheurs à comprendre les forces et les contraintes des données.

MOTS CLÉS : maillage de données; base de données administratives sur la santé; codes de diagnostic; études de validation; étude de synthèse