

Validity of Canadian discharge abstract data for hypertension and diabetes from 2002 to 2013

Jason Jiang MSc, Danielle Southern MSc, Cynthia A. Beck MD MSc, Matthew James MD PhD, Mingshan Lu PhD, Hude Quan MD PhD

Abstract

Background: Surveillance using coded administrative health data has shown that the prevalence of hypertension and diabetes in Canada increased substantially between 1998 to 2008. These findings require an assumption that the validity of hypertension and diabetes coding is stable over time. We tested this assumption by examining temporal trends in the validity of coding for hypertension and diabetes in the Canadian hospital Discharge Abstract Database.

Methods: We used the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database, a clinical registry, as the reference standard to evaluate the validity of the Discharge Abstract Database in recording hypertension and diabetes in Alberta. The APPROACH database contains data for all Alberta residents who have undergone cardiac catheterization and includes prospective ascertainment of comorbid conditions before each procedure. We linked patient data between the 2 databases for 2002 to 2013 using patient provincial health number. Temporal trends in sensitivity, specificity, positive predictive value, negative predictive value and Cohen κ were calculated for both hypertension and diabetes in the Discharge Abstract Database.

Results: We matched 63 483 patients between the APPROACH database and the Discharge Abstract Database. The validity of the Discharge Abstract Database for hypertension and diabetes remained mostly consistent over time. Between 2002 and 2013, sensitivity, specificity, positive predictive value and negative predictive value ranged from 66% to 87% for hypertension and from 81% to 98% for diabetes; the corresponding κ scores ranged from 0.50 to 0.62 and from 0.80 to 0.89. No significant differences in the validity of coding were found across age, sex or hospital location subgroups.

Interpretation: The validity of coding for hypertension and diabetes in the Discharge Abstract Database remained fairly consistent between 2002 and 2013. Our findings support the use of the Discharge Abstract Database for hypertension and diabetes surveillance in hospital settings.

Surveillance of chronic diseases has been established in many countries for monitoring disease burden and evaluating prevention and therapeutic programs. Temporal trends in disease prevalence are also used to forecast population health status and health resource needs. These trends inform health care policy such as resource allocation. The Public Health Agency of Canada created the Canadian Chronic Disease Surveillance System to provide governments and the public with new knowledge to inform efforts to reduce the risk of chronic conditions.¹ Collaborating with provincial and territorial health ministries, this system has successfully tracked hypertension and diabetes using administrative health data. According to the Canadian Chronic Disease Surveillance System, from 1998 to 2008, the prevalence of hypertension in Canada increased from 12.5% to 19.6%, and diabetes prevalence increased from 3.3% to 5.4%.^{1,2} The rate of increasing prevalence is alarming and has raised important questions regarding health policy and chronic disease prevention. However, before one interprets

these temporal trends in disease prevalence, potential errors from surveillance methods must be excluded or adjusted for.

A major question that remains unexamined is whether data validity has remained consistent over the period of this surveillance program. Although improving data quality over time is generally to be desired, it could be problematic for interpreting temporal trends in disease surveillance. For example, improving sensitivity of disease detection over time will result in increases in prevalence estimates that exceed actual values. The Canadian Chronic Disease Surveillance System monitors disease burden using diagnostic codes from the hospital Discharge Abstract

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Hude Quan, hquan@ucalgary.ca

CMAJ Open 2016. DOI:10.9778/cmajo.20160128

Database and the National Physician Database (physician claims).¹ The former contains information on all admissions at acute care facilities, and the latter captures billing data from physician visits. These 2 databases are attractive sources of information for disease surveillance because both are readily available and cover large populations in Canada.³

To assess whether this question has been addressed, we reviewed the literature. The published studies examined a relatively short period, usually conducting only 2 or 3 period comparisons over a maximum of 5 years.⁴⁻¹³ To fill this gap, we conducted a study to examine the validity of administrative codes from the Discharge Abstract Database for hypertension and diabetes between 2002 and 2013. We used a prospective clinical registry, the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database, as the reference standard. Our hypothesis was that hypertension and diabetes have been coded with increasing accuracy over time, which has contributed to their apparent rapid increased prevalence in the Canadian Chronic Disease Surveillance System.

Methods

Data sources

The Discharge Abstract Database contains information on all inpatient discharges in Alberta. After the patient is discharged, professional coders translate the diagnosis in medical charts into International Classification of Disease (ICD) codes.¹⁴ The ICD codes are then recorded in the Discharge Abstract Database and are used (e.g., by the Canadian Chronic Disease Surveillance System) in the calculation of chronic disease prevalence. Coders are primarily instructed to code conditions contributing to the admission.¹⁵ In Alberta, up to 50 conditions can be recorded with the use of ICD-10 codes for each admission.

The APPROACH database contains detailed clinical information, including comorbid conditions, for all patients who have undergone coronary catheterization in Alberta since 1995. Comorbidity data are collected directly by clinicians and catheterization laboratory staff before each procedure.¹⁶ Clinicians make a diagnosis based on the patient's presentation, symptoms, results of diagnostic tests and medical history as well as through direct interaction with the patient. The fact that patient information is collected prospectively in the clinical setting provides 3 major advantages. First, data collection follows a fixed structure and format. The APPROACH database contains a required form asking clinicians to enter information for a fixed set of comorbid conditions, whereas in the Discharge Abstract Database, coders are free to enter information on as many or as few conditions as they see fit. Second, the direct collection of data in the APPROACH database reduces the chance of error due to data translation, whereas in the Discharge Abstract Database, coders must collect and interpret information from medical charts that were written by clinicians. Third, the APPROACH database contains routine processes that check for data quality and completeness. Furthermore, procedures are in place to continuously standardize and improve measurement and capture of comorbid

conditions. These attributes of the APPROACH database make it an ideal choice for our reference standard.

Data linkage

We linked patient data from the APPROACH database with the Discharge Abstract Database from 2002 to 2013 using the unique identifier of provincial health number. The data unit for APPROACH is patient visits, but in the Discharge Abstract Database it is admissions. Thus, 1 APPROACH patient visit could be linked with multiple admissions. To avoid this, we selected a single index admission from the Discharge Abstract Database for each linked patient visit. For the index admission, we chose the admission for which the catheterization date fell between the admission and discharge dates. This means that data for a particular patient could potentially be in our analysis more than once if he or she underwent more than 1 cardiac catheterization procedure. We decided that this did not interfere with the intended objective of this study, since our goal was to assess the validity of the Discharge Abstract Database at multiple points in time and compare the differences between them.

Data from the Discharge Abstract Database were extracted by analysts at the Data Integration, Measurement and Reporting Department at Alberta Health Services. APPROACH data were provided by the APPROACH team. We used patient provincial health numbers from the APPROACH database to search for matching admissions in the Discharge Abstract Database. We then selected each APPROACH visit to a single Discharge Abstract Database admission that occurred during the time of that APPROACH visit.

Study variables

In the Discharge Abstract Database, hypertension was defined with ICD-10 codes I10.x, I11.x-I13.x and I15.x. Diabetes was defined with ICD-10 codes E10.x, E13.10, E13.12, E14.10, E14.12, E11.x, E13.0 and E14.0. We assessed hyperlipidemia, heart failure, cerebrovascular disease, peripheral vascular disease and chronic obstructive pulmonary disease as possible confounding comorbid conditions and defined them using previously validated codes.¹⁷ We used all 50 diagnostic fields when searching for the ICD-10 codes included in the coding definitions. In the APPROACH database, clinicians diagnose the presence or absence of these conditions at the time of catheterization following clinical guidelines and practice (Appendix 1, available at www.cmajopen.ca/content/4/4/E646/suppl/DC1).

We obtained information on age, sex and hospital location from the APPROACH database.

Data analysis

We used descriptive statistics to describe study populations. We calculated differences in the prevalence of hypertension and diabetes between 2002 and 2013, along with 95% confidence intervals, for both the APPROACH database and the Discharge Abstract Database. We calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and Cohen κ for Discharge Abstract Database recording of hypertension and diabetes for each year using APPROACH as the reference standard. We then stratified

these estimates by age, sex and hospital location to assess whether these factors significantly affected study outcomes.

Results

Of the 63 483 patients who were linked between the 2 databases, 70.5% were male; the average age in 2013 was 62.6 years

(Table 1). The prevalence of hypertension in 2013 was 57.4% in the APPROACH database and 57.8% in the Discharge Abstract Database; the prevalence of diabetes was 23.6% and 26.0%, respectively (Table 2). The differences were not statistically significant. For both hypertension and diabetes, the prevalence was similar for the 2 databases and across age, sex and hospital location over the entire study period (Figure 1, Table 2).

Table 1: Characteristics of patients who underwent coronary artery cauterization in Alberta in 2002 and 2013

Characteristic	Year; no. (%) of patients*				Absolute change between 2002 and 2013 (95% CI)	
	2002 (n = 4001)		2013 (n = 5580)		APPROACH	DAD
Age, mean ± SD, yr	63.8 ± 12.3		62.6 ± 14.8		-1.2 (-1.7 to -0.6)	
Men	2795 (69.9)		3935 (70.5)		0.6 (-1.2 to 2.5) 0.6%	
Location						
Calgary	1924 (48.1)		2300 (41.2)		-6.9 (-8.9 to -4.9)	
Edmonton	1706 (42.6)		2633 (47.2)		4.6 (2.5 to 6.6)	
Comorbid condition	APPROACH	DAD	APPROACH	DAD	APPROACH	DAD
Hyperlipidemia	2779 (69.4)	1912 (47.8)	3393 (60.8)	660 (11.8)	-8.7 (-10.6 to -6.8)	-36.0 (-37.7 to -34.2)
Heart failure	456 (11.4)	730 (18.2)	449 (8.0)	985 (17.6)	-3.3 (-4.6 to -2.2)	-0.6 (-2.2 to 1.0)
Cerebrovascular disease	234 (5.8)	150 (3.7)	253 (4.5)	98 (1.8)	-1.3 (-2.2 to -0.4)	-2.0 (-2.7 to -1.3)
Peripheral vascular disease	290 (7.2)	250 (6.2)	159 (2.8)	360 (6.4)	-4.4 (-5.3 to -3.5)	0.2 (-0.8 to 1.2)
Chronic obstructive pulmonary disease	552 (13.8)	434 (10.8)	800 (14.3)	512 (9.2)	0.6 (-0.9 to 1.9)	-1.7 (-2.9 to -0.4)

Note: APPROACH = Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease database, CI = confidence interval, DAD = Canadian hospital Discharge Abstract Database.
*Except where noted otherwise.

Table 2: Change in prevalence of hypertension and diabetes between 2002 and 2013

Condition; variable	Year; prevalence, %				Absolute change between 2002 and 2013 (95% CI)*	
	2002		2013		APPROACH	DAD
Hypertension	55.8	54.9	57.4	57.8	1.6 (-0.4 to 3.6)	2.9 (0.9 to 4.9)
Age, yr						
18-64	48.5	47.3	49.1	49.2	0.6 (-2.2 to 3.4)	1.9 (-0.9 to 4.7)
≥ 65	63.5	62.9	67.7	68.4	4.2 (1.4 to 7.0)	5.6 (2.8 to 8.4)
Sex						
Male	53.5	53.2	56.1	56.7	2.6 (0.2 to 5.0)	3.5 (1.1 to 5.9)
Female	61.1	58.9	60.4	60.4	-0.7 (-4.3 to 2.9)	1.5 (-2.2 to 5.1)
City						
Calgary	54.5	55.6	56.4	56.1	1.9 (-1.1 to 4.9)	0.6 (-2.4 to 3.6)
Edmonton	56.9	57.4	55.4	60.7	-1.5 (-4.5 to 1.5)	3.2 (0.2 to 6.2)
Diabetes	23.5	23.9	23.6	26.0	0.1 (-1.6 to 1.8)	2.1 (0.4 to 3.9)
Age, yr						
18-64	19.5	20.3	19.6	22.2	0.04 (-2.2 to 2.3)	1.9 (-0.4 to 4.2)
≥ 65	27.6	27.7	28.6	30.7	1.0 (-1.7 to 3.6)	3.1 (0.4 to 5.8)
Sex						
Male	22.6	23.0	22.8	25.7	0.2 (-1.8 to 2.2)	2.7 (0.6 to 4.7)
Female	25.4	26.0	25.3	26.8	0.02 (-3.3 to 3.2)	0.9 (-2.4 to 4.1)
City						
Calgary	24.0	24.4	22.8	26.4	-1.2 (-3.8 to 1.4)	2.0 (-0.6 to 4.6)
Edmonton	22.8	24.0	23.0	25.3	0.2 (-2.4 to 2.7)	1.3 (-1.4 to 3.9)

Note: APPROACH = Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease database, CI = confidence interval, DAD = Canadian hospital Discharge Abstract Database.
*Minor discrepancies are due to rounding.

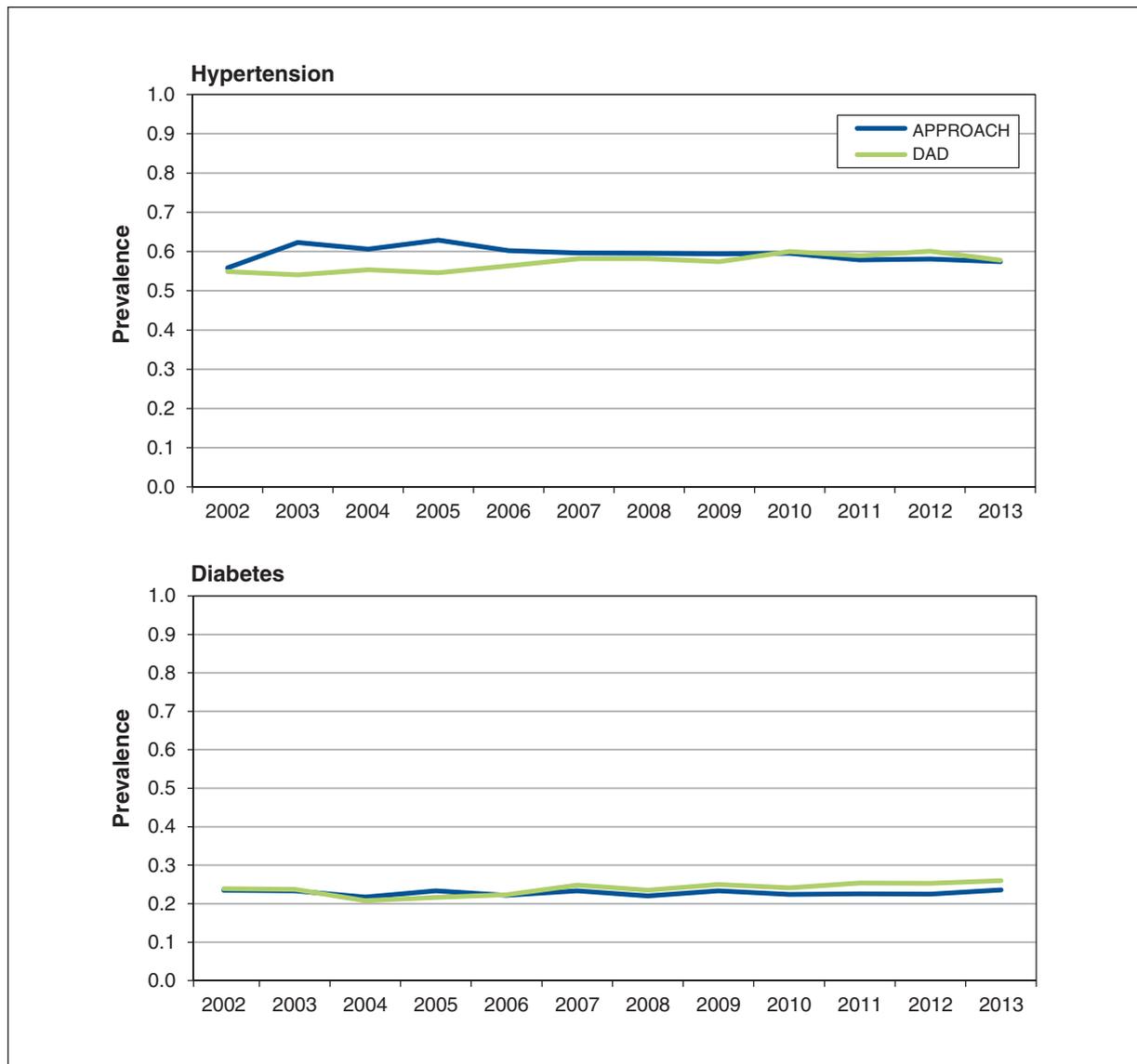


Figure 1: Prevalence of hypertension and diabetes in Alberta from 2002 to 2013. APPROACH = Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease database, DAD = Canadian hospital Discharge Abstract Database.

The validity of Discharge Abstract Database coding for hypertension and diabetes, with the APPROACH database as the reference standard, also remained consistent over time (Figure 2). For hypertension, sensitivity ranged from 74.6% to 82.5%, specificity from 71.0% to 80.9%, PPV from 79.3% to 87.0%, NPV from 65.5% to 78.0%, and κ scores from 0.50 to 0.62 (Table 3). Overall, the validity of hypertension coding in the Discharge Abstract Database remained mostly consistent from 2002 to 2013. For diabetes, sensitivity ranged from 81.5% to 92.1%, specificity from 93.9% to 97.0%, PPV from 81.4% to 90.5%, NPV from 94.5% to 97.6%, and κ scores from 0.80 to 0.89 (Table 4).

When the validation trends for hypertension and diabetes coding were stratified by age (18–64 yr v. 65 yr or older), sex (male v. female) and hospital location (Calgary v. Edmonton), there were no significant differences for either hypertension or diabetes (results not shown).

Interpretation

Our study of coding of hypertension and diabetes in the hospital Discharge Abstract Database in Alberta between 2002 and 2013 gave 3 main findings. First, hypertension and diabetes coding in this database provided prevalence estimates similar to those obtained from a prospective clinical registry, the APPROACH database. Second, current coding algorithms using the Discharge Abstract Database show high validity for diabetes and fair validity for hypertension, based on sensitivity, specificity, PPV, NPV and κ . Third, validity was fairly consistent over the 12-year study period and across age groups, sex and hospital location. These findings support the use of the Discharge Abstract Database for hypertension and diabetes surveillance and suggest that recently observed increases in the prevalence of these conditions are unlikely to have been due to improved data quality.

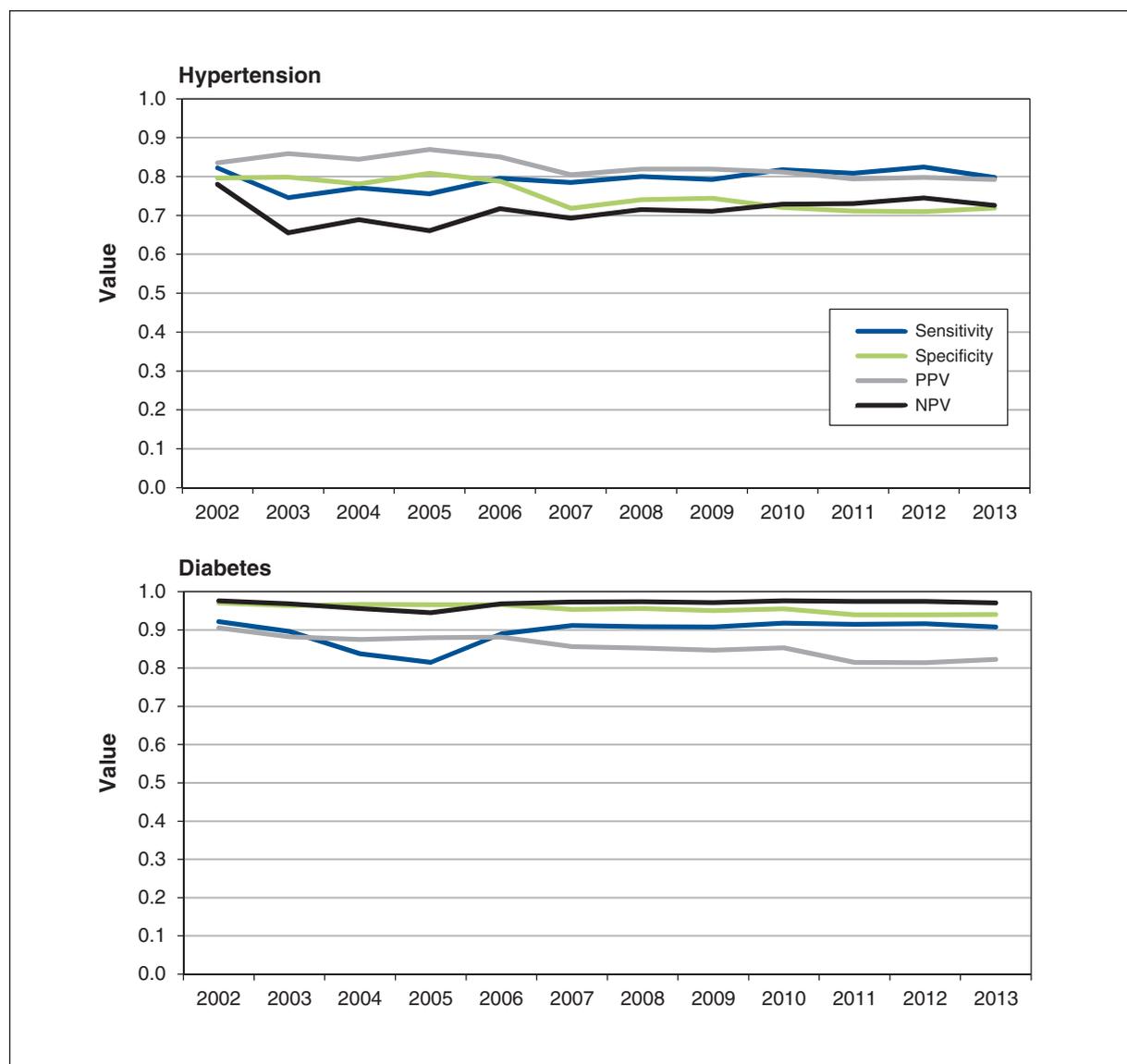


Figure 2: Validity of coding for hypertension and diabetes in the Discharge Abstract Database from 2002 to 2013, with the APPROACH database as a reference standard. NPV = negative predictive value, PPV = positive predictive value.

Our finding of fair to high validity of coding for hypertension and diabetes in administrative data is consistent with previous studies. Khokhar and colleagues¹⁸ reviewed 18 validation studies of administrative data for defining diabetes and found that sensitivity ranged from 51.78% to 100%, specificity from 88% to 100%, PPV from 21% to 99%, and NPV from 60.32% to 99.63%. They reported that validity varied depending on study population and administrative data source. Our results are also consistent with previous validation studies for hypertension.^{19–22} Quan and colleagues^{23,24} reported that administrative data coding for hypertension, based on a case definition of “2 physician claims within 2 years or 1 hospitalization,” had the highest validity, with sensitivity of 75%, specificity of 94%, PPV of 81% and NPV of 92%. The strong validity of coding for hypertension and diabetes may be related to a minimum 2-year professional coder training program, central management of coding practice and resource allocation to record department in Alberta.

Our findings support the use of the Discharge Abstract Database for surveillance and temporal trend analysis for hypertension and diabetes. We found that data quality remained consistent regardless of patient age, sex or hospital location. In Canada, hypertension and diabetes are mostly managed in outpatient clinics. Our validation study contained data only from hospital admissions. Therefore, data quality at outpatient sites over time is unknown. In Alberta, as in most other Canadian provinces/territories, physicians submit claims with the ICD diagnosis.²⁵ Cunningham and colleagues²⁶ evaluated the quality of physician visit data (including inpatients and outpatient visits) in Alberta and concluded that physicians submitted a variety of ICD codes for claims and that the face validity of diagnosis coded in physician claims is substantially high. In keeping with our findings, 2 Canadian studies (Quan and colleagues¹¹ and Chen and colleagues⁵) analyzed inpatient and outpatient administrative data validity in defining hypertension and diabetes and

did not find significant changes between 2001 and 2014. However, Januel and colleagues⁹ reported that inpatient data validity improved after ICD-10 implementation in Switzerland.

Chronic disease surveillance is influenced by data quality. Because of imperfect data, accumulation of false-positive and false-negative cases over time with the use of administrative data could result in biased estimation of hypertension and diabetes prevalence. Peng and colleagues²⁷ evaluated these impacts on surveillance using sophisticated statistical methods and showed that surveillance is less likely biased because false-positive and false-negative cases are balanced out over time.

Strengths and limitations

The major strength of this study is its scope. The APPROACH database is a prospective population-based clinical registry with a large number of observations over several years. This makes the assessment of time trends possible. However, our study is subject to limitations. First, it assumes that the registry is an accurate and suitable reference standard. Unfortunately, clinical diagnoses in the registry have not yet been validated. Like many studies using clinical reference standards, our study does not confirm the accuracy of the reference standard. Clinical definitions of hypertension and diabetes changed slightly during the

Table 3: Validity of coding for hypertension in the Discharge Abstract Database*

Year; variable	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	Negative predictive value, % (95% CI)	κ (95% CI)	
2002	82.2 (79.9–84.5)	79.6 (76.9–82.3)	83.6 (81.3–85.8)	78.0 (75.3–80.7)	0.62 (0.59–0.64)	
2003	74.6 (72.5–76.7)	79.8 (77.3–82.3)	85.9 (84.1–87.7)	65.5 (62.9–68.2)	0.52 (0.50–0.54)	
2004	77.1 (75.1–79.1)	78.1 (75.6–80.6)	84.4 (82.6–86.3)	68.9 (66.3–71.5)	0.54 (0.52–0.56)	
2005	75.5 (73.5–77.6)	80.9 (78.4–83.3)	87.0 (85.3–88.7)	66.1 (63.4–68.7)	0.54 (0.52–0.56)	
2006	79.5 (77.5–81.5)	78.8 (76.3–81.3)	85.0 (83.2–86.9)	71.7 (69.1–74.4)	0.57 (0.55–0.60)	
2007	78.5 (76.4–80.6)	71.9 (69.1–74.6)	80.5 (78.4–82.5)	69.3 (66.5–72.1)	0.50 (0.48–0.52)	
2008	80.0 (77.9–82.0)	74.0 (71.3–76.7)	81.9 (79.9–83.9)	71.5 (68.7–74.3)	0.54 (0.51–0.56)	
2009	79.3 (77.2–81.4)	74.5 (71.7–77.2)	81.9 (79.9–84.0)	71.1 (68.3–73.8)	0.53 (0.51–0.56)	
2010	81.8 (79.8–83.7)	72.1 (69.3–74.8)	81.2 (79.2–83.1)	72.9 (70.1–75.6)	0.54 (0.52–0.56)	
2011	80.8 (78.8–82.9)	71.2 (68.4–73.9)	79.4 (77.3–81.4)	73.0 (70.3–75.7)	0.52 (0.50–0.55)	
2012	82.5 (80.6–84.4)	71.0 (68.3–73.7)	79.8 (77.8–81.7)	74.5 (71.9–77.1)	0.54 (0.52–0.56)	
2013	79.8 (77.8–81.8)	71.9 (69.3–74.5)	79.3 (77.3–81.3)	72.5 (70.0–75.1)	0.52 (0.49–0.54)	
Age, yr	18–64	76.9 (73.9–79.9)	77.6 (74.6–80.5)	76.8 (73.8–79.8)	77.7 (74.7–80.6)	0.54 (0.52–0.57)
	≥ 65	82.4 (79.8–85.0)	60.8 (56.0–65.7)	81.5 (78.8–84.1)	62.3 (57.4–67.1)	0.43 (0.40–0.47)
Sex	Male	79.2 (76.8–81.6)	72.1 (69.1–75.2)	78.4 (76.0–80.9)	73.1 (70.1–76.1)	0.51 (0.49–0.54)
	Female	81.1 (77.6–84.6)	71.3 (66.3–76.3)	81.2 (77.7–84.6)	71.2 (66.2–76.1)	0.52 (0.48–0.57)
City	Calgary	76.3 (73.0–79.7)	70.1 (66.0–74.1)	76.8 (73.5–80.1)	69.6 (65.5–73.6)	0.46 (0.43–0.50)
	Edmonton	86.5 (84.0–89.0)	71.5 (67.8–75.2)	79.0 (76.2–81.9)	81.0 (77.6–84.4)	0.59 (0.55–0.62)

Note: CI = confidence interval.

*With the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease database as the reference standard.

study period. However, within the scope of this study we were unable to address whether these minor changes affected the prevalence of hypertension and diabetes in the APPROACH database. Diagnoses in APPROACH are made during the APPROACH visit based on the clinician's review of the patient's presentation and symptoms, clinical observation and measurement as well as direct communication with the patient. We assume that the APPROACH clinicians make diagnosis based on the clinical definitions at the time; however, we have no way to verify this, because that information was not available in our data. Nonetheless, the use of trained clinicians for data

entry, standardized definitions, and strategies for verifying data quality and completeness within the registry support its use as a clinically accurate reference standard. The prospective data entry process within the APPROACH database also avoids a major source of error that occurs as a result of time lag and translation. Second, we did not validate outpatient data over time. This means that our results may not be generalizable to an outpatient population. Third, we validated data only among patients who underwent coronary catheterization. PPV and NPV are influenced by disease prevalence, and the high prevalence of hypertension and diabetes in patients with suspected

Table 4: Validity of coding for diabetes in the Discharge Abstract Database*

Year; variable	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	Negative predictive value, % (95% CI)	κ (95% CI)	
2002	92.1 (89.7–94.6)	97.0 (96.2–97.9)	90.5 (87.8–93.1)	97.6 (96.8–98.4)	0.89 (0.87–0.90)	
2003	89.6 (87.2–92.0)	96.3 (95.5–97.2)	88.2 (85.6–90.7)	96.8 (96.0–97.6)	0.85 (0.84–0.87)	
2004	83.8 (80.8–86.7)	96.7 (95.9–97.4)	87.5 (84.7–90.2)	95.5 (94.7–96.4)	0.82 (0.80–0.84)	
2005	81.5 (78.5–84.5)	96.6 (95.8–97.4)	87.9 (85.3–90.6)	94.5 (93.5–95.4)	0.80 (0.78–0.82)	
2006	88.9 (86.4–91.5)	96.6 (95.8–97.4)	88.1 (85.5–90.7)	96.8 (96.1–97.6)	0.85 (0.84–0.87)	
2007	91.1 (88.8–93.5)	95.4 (94.4–96.3)	85.6 (82.9–88.4)	97.3 (96.5–98.0)	0.85 (0.83–0.86)	
2008	90.8 (88.4–93.3)	95.5 (94.6–96.5)	85.2 (82.3–88.1)	97.4 (96.6–98.1)	0.84 (0.83–0.86)	
2009	90.7 (88.3–93.1)	95.0 (94.0–96.0)	84.6 (81.8–87.5)	97.1 (96.3–97.9)	0.84 (0.82–0.85)	
2010	91.8 (89.5–94.0)	95.4 (94.5–96.4)	85.3 (82.5–88.1)	97.6 (96.9–98.3)	0.85 (0.83–0.87)	
2011	91.4 (89.1–93.7)	93.9 (92.9–95.0)	81.5 (78.5–84.5)	97.4 (96.7–98.1)	0.82 (0.80–0.84)	
2012	91.6 (89.4–93.8)	93.9 (92.9–95.0)	81.4 (78.5–84.3)	97.5 (96.8–98.2)	0.82 (0.80–0.84)	
2013	90.7 (88.5–93.0)	94.0 (93.0–95.0)	82.3 (79.5–85.1)	97.0 (96.3–97.8)	0.82 (0.80–0.84)	
Age, yr	18–64	89.4 (85.9–92.9)	94.2 (92.9–95.5)	78.9 (74.5–83.3)	97.3 (96.4–98.3)	0.80 (0.77–0.82)
	≥ 65	91.8 (88.9–94.7)	93.7 (92.1–95.3)	85.3 (81.8–88.9)	96.6 (95.4–97.8)	0.84 (0.81–0.86)
Sex	Male	91.8 (89.2–94.3)	93.9 (92.7–95.1)	81.7 (78.3–85.1)	97.5 (96.7–98.3)	0.82 (0.80–0.84)
	Female	88.5 (84.1–92.9)	94.1 (92.3–96.0)	83.7 (78.7–88.6)	96.0 (94.4–97.6)	0.81 (0.78–0.84)
City	Calgary	91.6 (88.2–95.0)	92.8 (91.1–94.6)	79.1 (74.5–83.7)	97.4 (96.3–98.5)	0.80 (0.77–0.83)
	Edmonton	91.7 (88.6–94.9)	94.5 (93.1–95.9)	83.3 (79.3–87.4)	97.5 (96.5–98.5)	0.83 (0.81–0.86)

Note: CI = confidence interval.

*With the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease database as the reference standard.

coronary artery disease may produce higher estimates of PPV and lower estimates of NPV than in the general population. Finally, our analysis was limited to data from Alberta. Data quality may vary between institutions and processes for data collection management in other settings.²⁸ Caution should be applied in generalizing our findings to other regions or sites with less rigorous processes for coding administrative data. Our study also contained a very specialized population, consisting only of patients with suspected heart disease. Thus, our results may not be generalizable to a general clinical population.

Conclusion

We found that the quality of the Discharge Abstract Database remained consistent between 2002 and 2013 for both hypertension and diabetes. This finding supports the use of administrative data for surveillance of these conditions. Further research on temporal trends in quality of physician claims data and ambulatory care databases needs to be conducted to confirm the effect of data quality from these sources on trends in chronic disease prevalence.

References

1. Report from the Canadian Chronic Disease Surveillance System: hypertension in Canada, 2010. Ottawa: Public Health Agency of Canada; 2010. Available: www.phac-aspc.gc.ca/cd-mc/cvd-mcv/cdss-snsmc-2010/2-1-eng.php (accessed 2015 Apr. 27).
2. Diabetes in Canada: facts and figures from a public health perspective. Ottawa: Public Health Agency of Canada; 2011.
3. Population health research data repository holdings. Winnipeg: University of Manitoba; 2011. Available: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=Hospital (accessed 2016 Oct. 21).
4. Allen VM, Dodds L, Spencer A, et al. Application of a national administrative case definition of pre-existing diabetes mellitus in pregnancy. *Chronic Dis Inj Can* 2012;32:113-20.
5. Chen G, Khan N, Walker R, et al. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010;89:189-95.
6. Daneshvar P, Forster A, Dervin G. Accuracy of administrative coding in identifying hip and knee primary replacements and revisions. *J Eval Clin Pract* 2012;18:555-9.
7. De Coster C, Li B, Quan H. Comparison and validity of procedures coded with ICD-9-CM and ICD-10-CA/CCI. *Med Care* 2008;46:627-34.
8. Hagen EM, Rekan T, Gilhus NE, et al. Diagnostic coding accuracy for traumatic spinal cord injuries. *Spinal Cord* 2009;47:367-71.
9. Januel JM, Luthi JC, Quan H, et al. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. *BMC Health Serv Res* 2011;11:194.
10. Jetté N, Reid AY, Quan H, et al. How accurate is ICD coding for epilepsy? *Epilepsia* 2010;51:62-9.
11. Quan H, Khan N, Hemmelgarn BR, et al.; Hypertension Outcome and Surveillance Team of the Canadian Hypertension Education Programs. Validation of a case definition to define hypertension using administrative data. *Hypertension* 2009;54:1423-8.
12. Quan H, Li B, Saunders LD, et al.; IMECCHI Investigators. Assessing the 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.
13. Reid AY, St Germaine-Smith C, Liu M, et al. Development and validation of a case definition for epilepsy for use with administrative health data. *Epilepsy Res* 2012;102:173-9.
14. Schneeweiss S, Avorn J. A review of uses of health care utilization database for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-37.
15. Canadian Coding Standards for Version 2015 ICD-10-CA and CCI. Ottawa: Canadian Institute for Health Information; 2015.
16. Ghali WA, Knudtson ML. Overview of the Alberta Provincial Project for Outcome assessment in Coronary Heart Disease. *Can J Cardiol* 2000;16:1225-30.
17. Southern DA, Norris CM, Quan H, et al.; APPROACH Investigators. An administrative data merging solution for dealing with missing data in a clinical registry: adaptation from ICD-9 to ICD-10. *BMC Med Res Methodol* 2008;8:1.
18. Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 2016;6:e009952.
19. Muhajarine N, Mustard C, Roos L, et al. Comparison of survey and physician claims data for detecting hypertension. *J Clin Epidemiol* 1997;50:711-8.
20. 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;39:1839-57.
21. Lix L, Yogendran M, Burchill C, et al. Defining and validating chronic diseases: an administrative data approach. Winnipeg: Manitoba Centre for Health Policy; 2006.
22. Tu K, Campbell NR, Chen ZL, et al. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18-26.
23. Quan H, McAlister F, Khan N. The many faces of hypertension in Canada. *Curr Opin Cardiol* 2014;29:354-9.
24. Hennessy DA, Quan H, Faris PD, et al. Do coder characteristics influence validity of ICD-10 hospital discharge data? *BMC Health Serv Res* 2010;10:99.
25. Lix LM, Walker R, Quan H, et al.; CHEP-ORTF Hypertension Outcome and Surveillance Team. Features of physician services databases in Canada. *Chronic Dis Inj Can* 2012;32:186-93.
26. Cunningham CT, Cai P, Topps D, et al. Mining rich health data from Canadian physician claims: features and face validity. *BMC Res Notes* 2014;7:682.
27. Peng M, Chen G, Lix LM, et al.; Hypertension Outcome and Surveillance Team. Refining hypertension surveillance to account for potentially misclassified cases. *PLoS One* 2015;10:e0119186.
28. Quan H, Smith M, Bartlett-Esquilant G, et al.; Hypertension Outcome and Surveillance Team. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. *Can J Cardiol* 2012;28:152-4.

Affiliations: Departments of Community Health Sciences (Jiang, Southern, Beck, James, Lu, Quan), Psychiatry (Beck), Medicine (James) and Economics (Lu), University of Calgary, Calgary, Alta.

Contributors: Hude Quan is the primary investigator and Jason Jiang's primary supervisor for this study. He contributed substantially to study conception and design, data analysis and interpretation, and drafting of the manuscript. Danielle Southern contributed substantially to drafting and revising the manuscript. Cynthia Beck, Matthew James and Mingshan Lu provided substantial input into study conception and design, and data analysis and interpretation. Jason Jiang carried out the primary data analysis, contributed substantially to data interpretation and drafted the manuscript. All of the authors approved the final version to be published and agreed to act as guarantors of the work.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/4/4/E646/suppl/DC1