

Validity of diagnoses of SARS-CoV-2 infection in Canadian administrative health data: a multiprovince, population-based cohort study

Lisa M. Lix PhD, Christel Renoux MD PhD, Carolina Moriello MSc, Ko Long Choi MSc, Colin R. Dormuth PhD, Anat Fisher MD PhD, Matthew Dahl BSc, Fangyun Wu MSc, Ayesha Asaf MPH, J. Michael Paterson MSc; for the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators

Abstract

Background: Accurate coding of diagnoses of SARS-CoV-2 infection in administrative data benefits population-based studies about the epidemiology, treatment and outcomes of COVID-19. We describe the validity of diagnoses of SARS-CoV-2 infection recorded in hospital discharge abstracts, emergency department records and outpatient physician service claims from 3 Canadian provinces.

Methods: In this cohort study, population-based inpatient, emergency department and outpatient records were linked to SARS-CoV-2 polymerase chain reaction (PCR; reference standard) test results from British Columbia, Manitoba and Ontario for Apr. 1, 2020, to Mar. 31, 2021. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of diagnoses of SARS-CoV-2 infection were estimated for each quarter in the study period, overall and by province, age group and sex.

Results: Our study encompassed more than 13 million SARS-CoV-2 PCR test results. Specificity and NPV of diagnoses of SARS-CoV-2 infection were consistently high (i.e., most estimates were > 95%). Overall sensitivity estimates were 86.2%, 60.4% and 20.3% in the first quarter for inpatient, emergency department and outpatient cohorts, and 66.2%, 47.5% and 25.0% in the last quarter, respectively. For inpatients, overall PPV estimates ranged from 50.0% to 66.4%. For emergency department patients, overall PPV estimates were 76.9% and 68.3% in the first and last quarters, respectively. For outpatients, PPV estimates were 6.8% and 29.1% in the first and last quarters, respectively.

Interpretation: We found variations in the validity of diagnoses for SARS-CoV-2 infection recorded in different health care settings, geographic areas and over time. Our multiprovince validation study provides evidence about the potential use of inpatient and emergency department records as an alternative to population-based laboratory data for identification of patients with SARS-CoV-2 infection, but does not support the use of outpatient claims for this purpose.

Administrative health data are increasingly used to examine outcomes, risk factors and treatments at the population level for individuals diagnosed with SARS-CoV-2 infection. However, the validity of diagnosis coding in these data is largely unknown outside of inpatient settings.¹⁻⁵ Validity may vary between jurisdictions and over time owing to variations in testing protocols, diagnosis coding procedures and staff training. There is a substantial history of conducting diagnostic validation studies for administrative health data in Canada to inform studies about the validity of diagnoses of SARS-CoV-2 infection.^{6,7}

The Canadian Institute for Health Information published its first guidance about the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada* (ICD-10-CA) diagnosis codes for COVID-19 patients admitted to hospitals and emergency departments in March 2020.⁸

Provincial and territorial guidance for diagnosis and fee codes for physician service claims were also published. For example, Ontario introduced a COVID-19 diagnosis code in March 2020 to be used “when treating patients with suspected or confirmed COVID-19 and/or when treating a patient by telephone/video for suspected or confirmed COVID-19,”⁹ and British Columbia published COVID-19 codes for “services directly related to COVID-19.”¹⁰

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Lisa Lix, lisa.lix@umanitoba.ca

CMAJ Open 2023 September 5. DOI:10.9778/cmajo.20220152

To our knowledge, only Wu and colleagues³ have studied the validity of COVID-19 diagnoses in Canadian inpatient and emergency department records, and they focused on Alberta. Information about the validity of diagnoses of SARS-CoV-2 infection in administrative data is important to understand COVID-19 epidemiology, treatments and outcomes. Our study assessed the validity of diagnoses of SARS-CoV-2 infection recorded in inpatient hospital discharge abstracts, emergency department records and outpatient physician service claims in 3 Canadian provinces.

Methods

We undertook a population-based cohort study using administrative health data linked with SARS-CoV-2 polymerase chain reaction (PCR) laboratory test results from BC, Manitoba and Ontario (provinces for which we had access to linked data sources) for Apr. 1, 2020, through Mar. 31, 2021. Throughout the study period there was publicly available PCR testing in these provinces. We adopted the Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines with the Reporting of Studies Conducted Using Observational Routinely-collected Data extension.^{11,12}

Data sources

Four administrative health databases were used in each province: health insurance registry, physician service claims, emergency department discharge records and inpatient hospital discharge records (Appendix 1, Table S1, available at www.cmajopen.ca/content/11/5/E790/suppl/DC1). All data sources were linked at the individual level using anonymized health insurance numbers; only individuals with valid health insurance numbers were included. Health insurance registry files capture start and end dates of health insurance coverage, including loss of coverage due to death or migration; demographic and residence location information is also captured. Physician service claims capture services provided by specialists and general practitioners, including type and date of service, fee code and at least 1 diagnosis code associated with the reason for the service. Diagnosis codes are recorded using modifications of the eighth (Ontario) and ninth revisions of ICD (ICD-8 and ICD-9).¹³ We included claims for office, telephone and virtual patient consultations, home visits and long-term care visits. Emergency department discharge records contain information about visits to hospital-based emergency departments, including visit date, chief complaint and diagnoses, typically coded using ICD-10-CA. Inpatient hospital discharge records contain diagnostic and procedural information for acute hospital stays, including up to 25 diagnoses coded using ICD-10-CA.

SARS-CoV-2 laboratory test results were linked to administrative data. Test results were from the BC Ministry of Health COVID-19 Test Laboratory Data, the Ontario Laboratories Information System, and the Manitoba Cadham Provincial COVID-19 Laboratory Testing and Results Database.

Study cohorts

Inpatient, emergency department and outpatient study cohorts were composed of provincial health insurance registrants with at least 1 inpatient hospital discharge record, emergency department discharge record and outpatient physician service claim, respectively. Cohort exclusions were based on missing sex or health care coverage, and overlap in dates of hospitalization, emergency department visits and outpatient visits. The cohorts were stratified into 3-month (quarter) subgroups.

Study measures

Positive test cases, negative test cases and no-test cases were identified from laboratory data. A positive test case had at least 1 positive PCR test with a specimen collection date within the quarter, and a negative test case had at least 1 negative PCR test with a specimen collection date and no positive PCR tests within the quarter. No-test cases had no PCR tests, or only indeterminate PCR tests within the quarter.

The ICD-10-CA code U07.1 in any position was used to ascertain diagnosed cases in hospital and emergency department records (Appendix 1, Table S2). For the inpatient cohort, true positive cases had a COVID-19 diagnosis code and specimen collection date for a positive PCR test between (and including) hospital admission and discharge dates. For the emergency department cohort, true positive cases had a COVID-19 diagnosis code with a service date at most 2 days before the specimen collection date or 2 days after the specimen collection date for a positive PCR test.

For the outpatient cohort, newly developed COVID-19 coding directives provided by each ministry of health was used for case ascertainment (Appendix 1, Table S2). In Ontario and Manitoba, the diagnosis codes were 080⁹ and 079.82,¹⁴ respectively. In BC, case ascertainment was initially based on the newly developed diagnosis code C19.¹⁰ However, this diagnosis code was not associated with any claims during the study period; accordingly, newly developed service codes relevant to COVID-19 were used for case ascertainment.¹⁰ True positive cases had a COVID-19 diagnosis code with a service date in physician claims at most 2 days before the specimen collection date or 2 days after the specimen collection date for a positive PCR test.

Statistical analysis

We used frequencies, percentages, means and standard deviations (SDs) to describe the cohort (i.e., age group, sex, income quintile, and rural or urban residence). Area-level income quintile was based on postal code of residence and household income from the Statistics Canada census.^{15,16} Rural or urban residence was based on each province's definition; for example, Manitoba rural residents lived outside of the major urban centres of Winnipeg and Brandon.

Validity was assessed using sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). We reported the estimates as percentages with 95%

confidence intervals (CIs). The Youden Index (calculated as sensitivity + specificity - 1) was also reported.¹⁷ We produced validation measure estimates for each cohort and quarter for the 3 provinces combined and then separately for each province. Estimates were also stratified by sex and age group (< 65 yr, 65–79 yr, ≥ 80 yr).

To assess the robustness of our findings, we performed 4 prespecified sensitivity analyses using larger case ascertainment windows as per previous studies.^{1,2} In the first, for the inpatient cohort, true positive cases had a specimen collection date for a positive test in the period from 7 days before the admission date to the discharge date. In the second, for the inpatient cohort, the time window extended from 14 days before to 14 days after the hospital admission date to ascertain true positive cases, consistent with the work of Kluberg and colleagues.^{1,2} In the third and fourth sensitivity analyses, for the emergency department and outpatient cohorts, respectively, we identified true positives used a time window from 5 days before to 5 days after the specimen collection date for a positive test.

Ethics approval

Approvals were provided by the following ethics boards: the Clinical Research Ethics Board at the University of British Columbia and the Health Research Ethics Board at the University of Manitoba.

Results

The study cohorts were composed of approximately 1.3 million inpatients, 3.2 million emergency department patients and 15.1 million outpatients (Figure 1). Ontario residents accounted for 51.2%, 77.9% and 67.3% of the inpatient, emergency department and outpatient cohorts, respectively. Average age was 55.9 years for the inpatient cohort, 43.3 years for the emergency department cohort and 44.4 years for the outpatient cohort (Table 1; Appendix 1, Tables S3–S5).

This study encompassed more than 13 million SARS-CoV-2 PCR test results. Monthly PCR tests, positive PCR tests and PCR test rates per 10 000 population are reported in Appendix 1, Table S6. The percentage of positive tests ranged from 0.1% to 12.6%, and positive test rates per 10 000 population ranged from 1 to 51 per month.

Figures 2, 3 and 4 contain overall estimates of sensitivity, specificity, PPV and NPV for the inpatient, emergency department and outpatient cohorts, respectively, by quarter. Province-specific estimates are reported in Tables 2, 3 and 4; case frequencies are in Appendix 1, Tables S7 to S9. Specificity and NPV estimates were consistently high and frequently exceeded 95%.

For the inpatient cohort, the overall sensitivity was 86.2% (95% CI 84.2%–88.1%) in the first quarter (i.e., Q1); it dropped to 66.2% (95% CI 64.7%–67.6%) in the last quarter

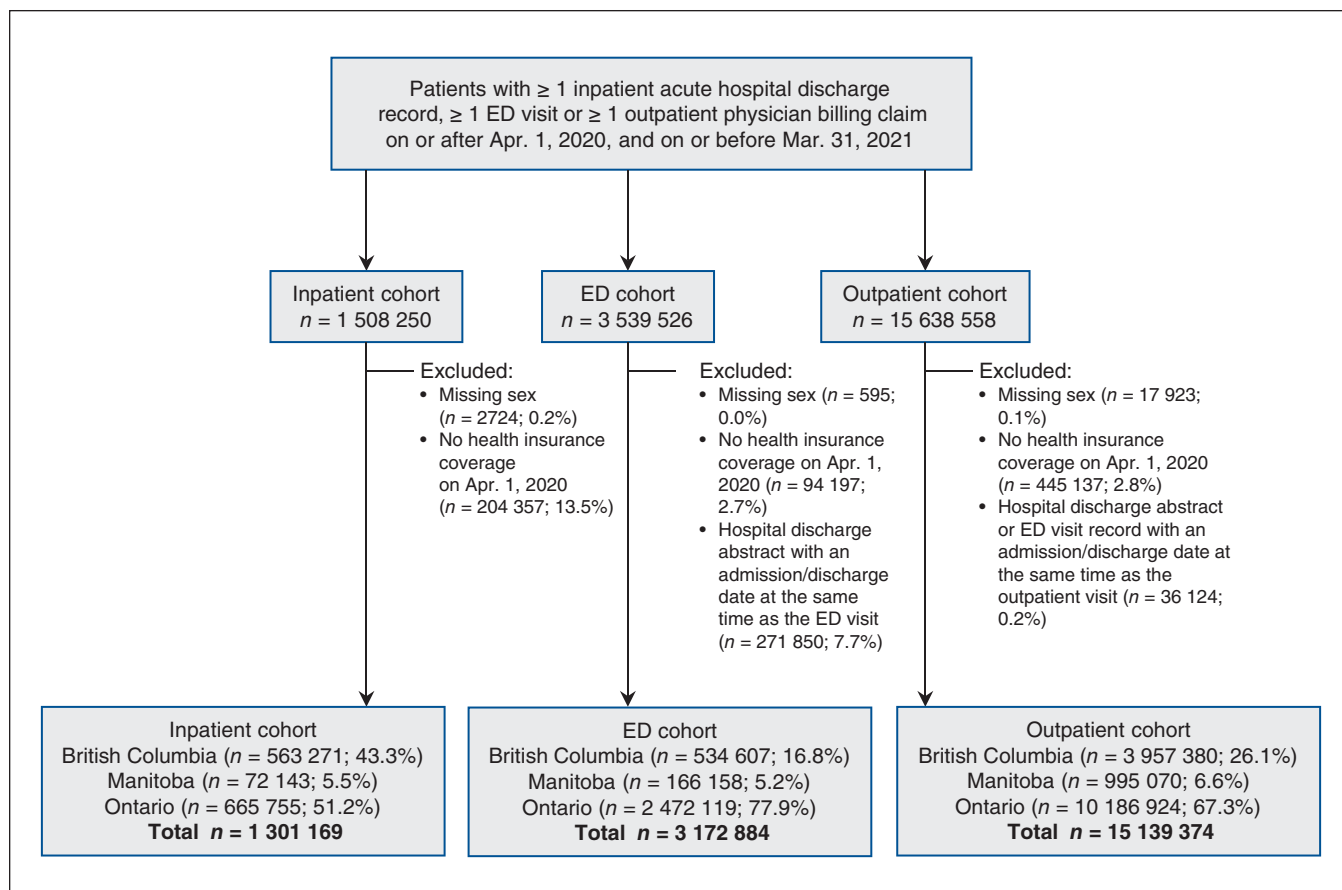


Figure 1: Flow diagram for construction of study cohorts. Note: ED = emergency department.

Table 1: Characteristics of inpatient, emergency department and outpatient cohorts			
Characteristic	No. (%)*		
	Inpatient <i>n</i> = 1 301 169	Emergency department <i>n</i> = 3 172 884	Outpatient <i>n</i> = 15 139 374
Age, yr			
Mean ± SD	55.9 ± 22.3	43.3 ± 23.5	44.4 ± 23.3
< 65	752 901 (57.9)	2 496 872 (78.7)	11 752 789 (77.6)
65–79	352 603 (27.1)	464 006 (14.6)	2 506 922 (16.6)
≥ 80	195 665 (15.0)	212 006 (6.7)	879 663 (5.8)
Sex			
Male	556 729 (42.8)	1 521 172 (47.9)	6 926 694 (45.8)
Female	744 440 (57.2)	1 651 712 (52.1)	8 212 680 (54.2)
Income quintile			
1st (lowest)	246 685 (19.0)	639 404 (20.1)	2 463 368 (16.3)
2nd	197 114 (15.2)	559 575 (17.6)	2 370 653 (15.7)
3rd	186 366 (14.3)	549 125 (17.3)	2 460 442 (16.3)
4th	181 575 (13.9)	540 065 (17.0)	2 523 160 (16.7)
5th (highest)	253 689 (19.5)	570 187 (17.9)	3 043 805 (20.1)
Missing	235 740 (18.1)	314 528 (9.9)	2 277 946 (15.1)
Area of residence			
Rural	184 737 (14.2)	442 454 (13.9)	1 690 020 (11.2)
Urban	1 113 889 (85.6)	2 721 432 (85.8)	13 412 772 (88.6)
Missing	2543 (0.2)	8998 (0.3)	36 582 (0.2)

Note: SD = standard deviation.
*Unless stated otherwise.

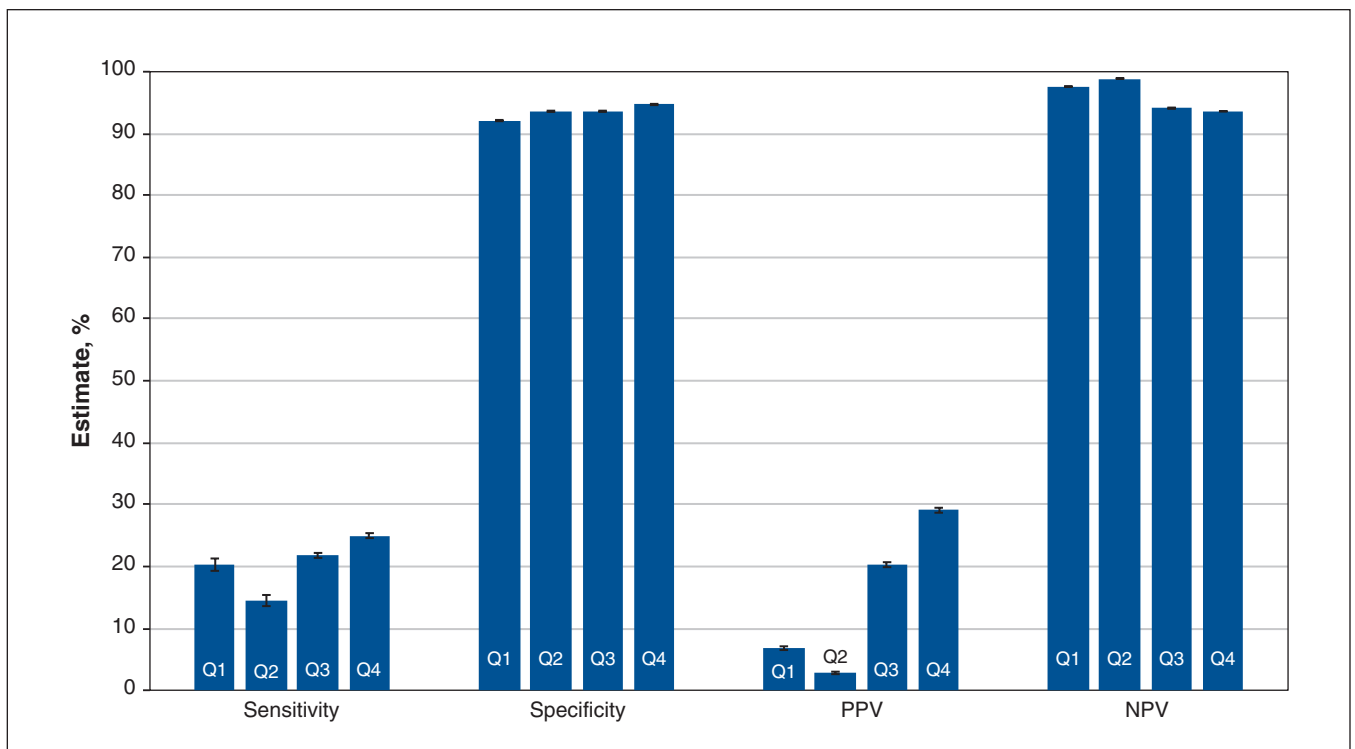


Figure 2: Overall validation estimates by quarter (Q): inpatient cohort. Note: NPV = negative predictive value, PPV = positive predictive value, Q1 = April 2020–June 2020, Q2 = July 2020–September 2020, Q3 = October 2020–December 2020, Q4 = January 2021–March 2021. Error bars represent 95% confidence intervals.

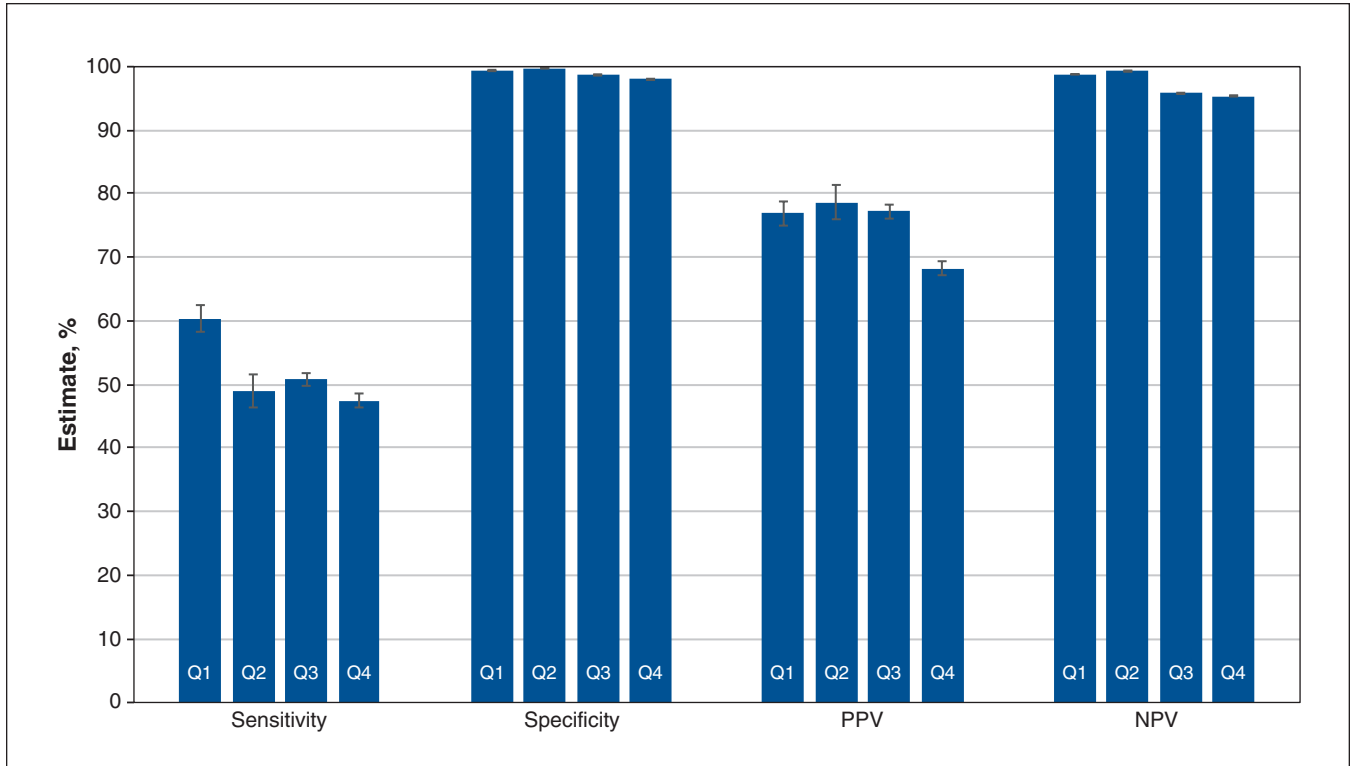


Figure 3: Overall validation estimates by quarter (Q): emergency department cohort. Note: NPV = negative predictive value, PPV = positive predictive value, Q1 = April 2020–June 2020, Q2 = July 2020–September 2020, Q3 = October 2020–December 2020, Q4 = January 2021–March 2021. Error bars represent 95% confidence intervals.

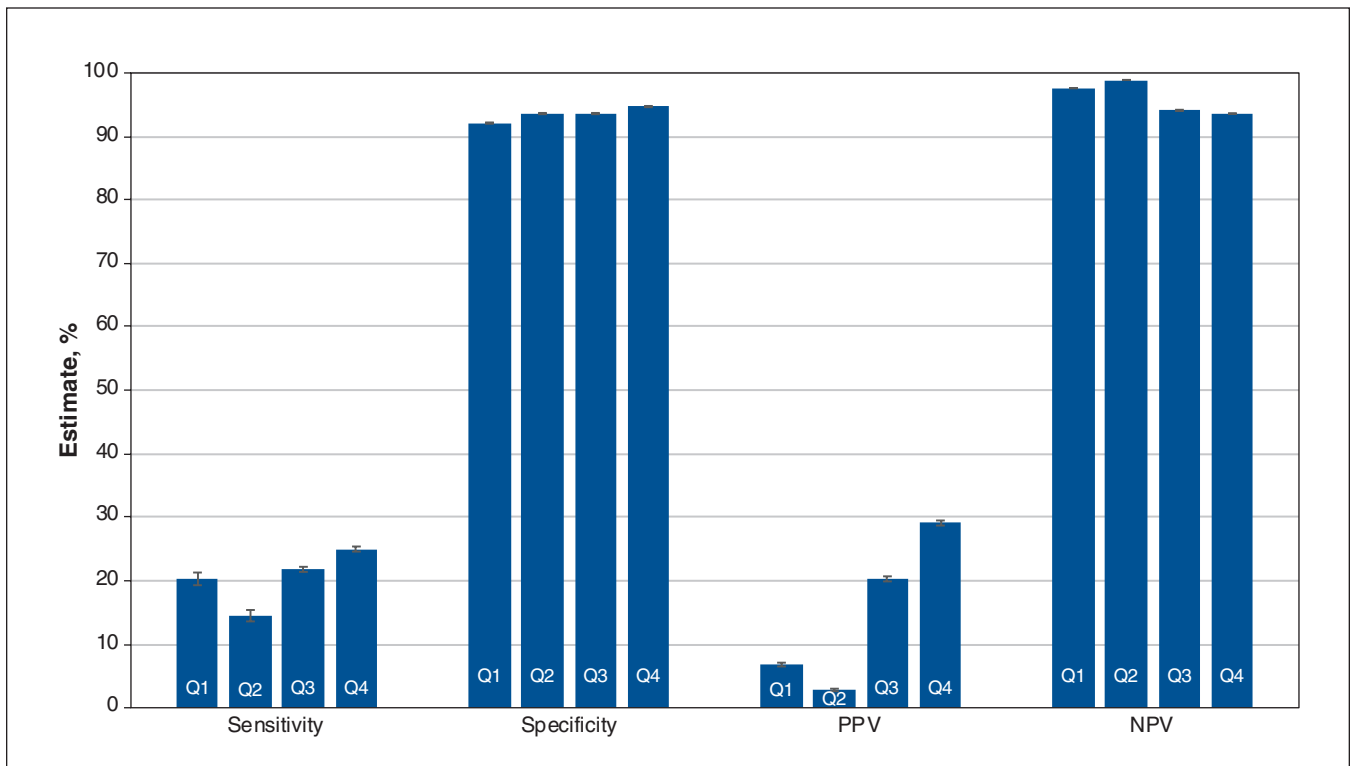


Figure 4: Overall validation estimates by quarter (Q): outpatient cohort. Note: NPV = negative predictive value, PPV = positive predictive value, Q1 = April 2020–June 2020, Q2 = July 2020–September 2020, Q3 = October 2020–December 2020, Q4 = January 2021–March 2021. Error bars represent 95% confidence intervals.

Table 2: Validation estimates by province and quarter: inpatient cohort

Measure	Estimate, % (95% CI)			
	Q1 April 2020–June 2020	Q2 July 2020–September 2020	Q3 October 2020–December 2020	Q4 January 2021–March 2021
British Columbia				
Sensitivity	88.4 (84.8–92.1)	56.1 (49.8–62.4)	52.0 (49.7–54.3)	53.5 (51.6–55.4)
Specificity	97.5 (97.3–97.7)	97.7 (97.5–97.8)	97.5 (97.4–97.7)	98.3 (98.2–98.5)
PPV	30.0 (26.9–33.0)	17.8 (15.1–20.6)	46.9 (44.8–49.1)	58.9 (57.0–60.9)
NPV	99.9 (99.8–99.9)	99.6 (99.5–99.7)	98.0 (97.8–98.1)	97.9 (97.8–98.1)
Youden Index	0.86	0.54	0.49	0.52
Manitoba				
Sensitivity	88.9 (77.0–100.0)	75.4 (67.5–83.3)	59.1 (56.4–61.8)	48.2 (44.1–52.3)
Specificity	99.6 (99.4–99.7)	99.3 (99.1–99.5)	95.3 (94.9–95.6)	98.5 (98.3–98.7)
PPV	46.2 (32.6–59.7)	60.6 (52.5–68.6)	55.4 (52.7–58.0)	53.6 (49.4–57.9)
NPV	99.9 (99.9–100.0)	99.6 (99.5–99.8)	95.9 (95.5–96.2)	98.1 (97.9–98.3)
Youden Index	0.88	0.75	0.54	0.47
Ontario				
Sensitivity	86.0 (84.8–87.1)	73.0 (70.2–75.8)	62.3 (61.0–63.6)	71.0 (70.1–72.0)
Specificity	98.8 (98.8–98.9)	99.8 (99.8–99.8)	98.8 (98.8–98.9)	98.0 (97.9–98.1)
PPV	75.0 (73.6–76.4)	73.6 (70.9–76.4)	68.3 (66.9–69.6)	68.9 (67.9–69.8)
NPV	99.4 (99.4–99.5)	99.8 (99.7–99.8)	98.5 (98.4–98.5)	98.2 (98.1–98.3)
Youden Index	0.85	0.73	0.61	0.69

Note: CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value, Q = quarter.

Table 3: Validation estimates by province and quarter: emergency department cohort

Measure	Estimate, % (95% CI)			
	Q1 April 2020–June 2020	Q2 July 2020–September 2020	Q3 October 2020–December 2020	Q4 January 2021–March 2021
British Columbia				
Sensitivity	52.4 (46.4–58.4)	44.3 (40.8–47.9)	31.8 (30.3–33.2)	28.7 (27.2–30.2)
Specificity	99.8 (99.8–99.9)	99.7 (99.7–99.8)	98.2 (98.1–98.3)	98.1 (98.0–98.3)
PPV	79.1 (73.1–85.1)	80.0 (76.2–83.8)	61.4 (59.2–63.6)	58.1 (55.8–60.4)
NPV	99.4 (99.3–99.5)	98.6 (98.4–98.7)	94.0 (93.8–94.3)	93.8 (93.6–94.1)
Youden Index	0.52	0.44	0.30	0.27
Manitoba				
Sensitivity	0 (0–21.9)	2.6 (0–5.18)	9.9 (8.6–11.1)	11.6 (9.5–13.6)
Specificity	99.9 (99.8–99.9)	99.8 (99.7–99.9)	97.3 (97.0–97.5)	98.9 (98.7–99.0)
PPV	0 (0–37.1)	12.1 (1.0–23.3)	34.2 (30.6–37.9)	39.9 (34.1–45.7)
NPV	99.8 (99.8–99.8)	98.9 (98.7–99.1)	88.3 (87.8–88.8)	94.5 (94.2–94.9)
Youden Index	0	0.02	0.07	0.11
Ontario				
Sensitivity	61.0 (59.7–62.2)	52.3 (50.6–54.1)	59.0 (58.4–59.7)	52.3 (51.6–53.0)
Specificity	99.4 (99.3–99.4)	99.8 (99.8–99.9)	98.9 (98.8–98.9)	97.9 (97.9–98.0)
PPV	77.0 (75.8–78.2)	79.6 (77.8–81.3)	81.3 (80.7–82.0)	70.0 (69.2–70.7)
NPV	98.6 (98.6–98.7)	99.4 (99.4–99.5)	96.6 (95.6–96.7)	95.7 (95.6–95.8)
Youden Index	0.60	0.52	0.58	0.50

Note: CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value, Q = quarter.

Table 4: Validation estimates by province and quarter: outpatient cohort

Measure	Estimate, % (95% CI)			
	Q1 April 2020–June 2020	Q2 July 2020–September 2020	Q3 October 2020–December 2020	Q4 January 2021–March 2021
British Columbia				
Sensitivity	10.2 (8.6–11.8)	6.1 (5.4–6.7)	2.4 (2.2–2.5)	2.5 (2.3–2.6)
Specificity	88.5 (88.3–88.7)	91.7 (91.6–91.8)	96.9 (96.8–96.9)	98.2 (98.2–98.3)
PPV	1.0 (0.9–1.2)	1.5 (1.4–1.7)	5.6 (5.2–5.9)	12.9 (12.2–13.7)
NPV	98.8 (98.8–98.9)	97.9 (97.8–97.9)	92.7 (92.6–92.7)	90.5 (90.4–90.6)
Youden Index	–0.01	–0.02	–0.01	–0.01
Manitoba				
Sensitivity	1.3 (0–3.8)	1.7 (0.84–2.5)	5.0 (4.5–5.4)	6.3 (5.5–7.2)
Specificity	99.5 (99.4–99.5)	99.7 (99.7–99.8)	96.7 (96.5–96.8)	95.2 (95.0–95.3)
PPV	0.6 (0–1.7)	8.3 (4.3–12.4)	14.0 (12.8–15.1)	5.7 (5.0–6.5)
NPV	99.8 (99.7–99.8)	98.7 (98.6–98.7)	90.3 (90.1–90.4)	95.6 (95.5–95.8)
Youden Index	0.01	0.01	0.02	0.02
Ontario				
Sensitivity	21.1 (20.6–21.7)	20.1 (19.3–20.9)	33.0 (32.7–33.3)	35.6 (35.3–35.9)
Specificity	92.5 (92.4–92.5)	93.7 (93.7–93.8)	92.0 (92.0–92.1)	93.5 (93.4–93.5)
PPV	8.5 (8.2–8.7)	3.2 (3.1–3.3)	22.5 (22.2–22.7)	31.1 (30.8–31.4)
NPV	97.3 (97.2–97.3)	99.1 (99.1–99.2)	95.2 (95.1–95.2)	94.6 (94.6–94.6)
Youden Index	0.14	0.14	0.25	0.29

Note: CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value, Q = quarter.

(i.e., Q4). The overall PPVs were 66.4% (95% CI 64.5%–68.4%) in Q1 and 66.3% (95% CI 65.0%–67.6%) in Q4. The lowest overall PPV was 50.0% (95% CI 46.8%–53.2%) in Q2. Province-specific PPVs for Q1 ranged from 30.0% (95% CI 26.9%–33.0%) in BC to 75.0% (95% CI 73.6%–76.4%) in Ontario. In BC, the PPV was low in Q1 and Q2, then increased. In Ontario, the PPV dropped in Q3 and Q4 compared with Q2. The Youden Index ranged from 0.49 to 0.88.

For the emergency department cohort, the overall sensitivity was 60.4% (95% CI 58.3%–62.5%) in Q1 and 47.5% (95% CI 46.5%–48.6%) in Q4. The overall PPVs were 76.9% (95% CI 75.0%–78.8%) and 68.3% (95% CI 67.2%–69.4%) in Q1 and Q4, respectively. Sensitivity was poor for Manitoba throughout the study period, with a maximum of 11.6% (95% CI 9.5%–13.6%) in Q4; the corresponding PPV in Q4 was 39.9% (95% CI 34.1%–45.7%). In comparison, maximum sensitivity for Ontario was 61.0% (95% CI 59.7%–62.2%) in Q1 and the PPV was 77.0 (95% CI 75.8%–78.2%). The Youden Index had a maximum of 0.60.

For the outpatient cohort, overall sensitivity was 20.3% (95% CI 19.4%–21.3%) in Q1 and 25.0% (95% CI 24.6%–25.4%) in Q4. The overall PPVs were 6.8% (95% CI 6.5%–7.1%) in Q1 and 29.1% (95% CI 28.7%–29.5%) in Q4. Sensitivity in Manitoba and Ontario was low but increased slightly from 1.3% (95% CI 0%–3.8%) and 21.1% (95% CI 20.6%–21.7%), respectively, in Q1, to 6.3% (95% CI 5.5%–

7.2%) and 35.6% (95% CI 35.3%–35.9%), respectively, in Q4. In BC, sensitivity declined, falling from 10.2% (95% CI 8.6%–11.8%) in Q1 to 2.5% (95% CI 2.3%–2.6%) in Q4. The PPV increased slightly from 1.0% (95% CI 0.9%–1.2%) in Q1 to 12.9% (95% CI 12.2%–13.7%) in Q4. The Youden Index had a maximum of 0.29.

Overall sensitivity and PPV generally increased across age groups in the inpatient cohort, but declined across age groups in the emergency department and outpatient cohorts (Appendix 1, Tables S10–S12). No consistent pattern was observed for sex (Appendix 1, Tables S10–S12).

For the predefined sensitivity analyses for the inpatient cohort (Appendix 1, Table S13 and Table S14), expanding the duration of the case ascertainment window led to absolute increases in estimated sensitivity of up to 46%, with the largest in BC. The first sensitivity analysis resulted in greater improvements in overall sensitivity and PPV estimates than the second; temporal trends were similar to those observed in the primary analysis. For the third sensitivity analysis (i.e., emergency department cohort; Appendix 1, Table S15), expanding the case ascertainment window resulted in absolute increases of up to 10% in sensitivity. Increases in PPV were small. For the outpatient cohort, expanding the case ascertainment window led to absolute increases in sensitivity and PPV of up to 18%, although most increases were small (i.e., less than 5%; Appendix 1, Table S16).

Interpretation

Our multiprovince validation study showed that diagnoses of SARS-CoV-2 infection had PPV estimates ranging from 17.8% to 75.0% in inpatient records, 0.0% to 81.3% in emergency department records, and 0.6% to 31.1% in outpatient records, illustrating substantial variation by province and over time. Positive predictive value estimates either improved or remained stable over time, and sensitivity estimates generally declined. Positive predictive value estimates are dependent on prevalence of positive tests, which varied considerably. Expanding the duration of the observation window for ascertaining diagnoses of SARS-CoV-2 infection in health care records improved sensitivity and PPV estimates for inpatient data, but the effect for outpatient data was generally small.

Using US inpatient data from May to October 2020, Kluberger and colleagues^{1,2} reported sensitivity estimates of 95% and PPV estimates of 81% using ICD-10 code U07.1. Similarly, Kadri and colleagues⁴ reported sensitivity estimates of 98% and PPV estimates of 92% from April to May 2020. Estimates from our study were lower in all provinces. However, PPVs improved over time in both Manitoba and BC and decreased only slightly in Ontario.

Our study provides important information about the validity of COVID-19 diagnosis coding in emergency department records and outpatient physician claims. For the emergency department cohort, overall validity was poorer than in a prior Canadian study.³ Positive predictive value estimates were highest in Ontario. Performance was noticeably poorer in Manitoba than in the other 2 provinces. Manitoba does not use the National Ambulatory Care Reporting System for emergency department records and has fewer fields for diagnosis codes in emergency department records than Ontario and BC, which may account in part for this finding.

The validity of COVID-19 diagnosis coding in outpatient claims was poor and only minimally affected by expanding the case ascertainment window. This finding may be attributed to limited access to family physicians, particularly during the early months of the pandemic, the multiple reasons a person may consult their physician regarding COVID-19, the likelihood that testing-related visits were directed to hospital- or community-based testing clinics rather than doctors' offices, and the time needed for physicians and billing clerks to become accustomed to new diagnosis or fee codes. Our findings do not support use of physician service claims as a substitute for population-based laboratory data to identify patients with SARS-CoV-2 infection.

Strengths of our study include assessment of diagnostic validity in inpatient and outpatient settings, during multiple periods and in 3 provinces. Our access to population-wide, community-based PCR laboratory test results made validation possible outside of inpatient hospital settings.

Limitations

The generalizability of our findings to outpatient physician service claims from other provinces and territories is unknown; each jurisdiction implemented its own COVID-19

coding for outpatient claims. In addition, while SARS-CoV-2 PCR laboratory testing was openly and widely accessible to symptomatic patients throughout the study period, the results of these PCR tests could influence diagnosis coding behaviour. Our findings may not generalize to settings where SARS-CoV-2 PCR testing policies and practice differed from those in these provinces during the study period, although many countries appear to have implemented similar testing policies.¹⁸ Another limitation is the potential lack of independence between PCR testing and the COVID-19 diagnosis. For the 2 to be independent, a positive or negative test result would need to be possible for all cohort members and in all settings. For example, we found that in Ontario, 47% of inpatient hospitalizations were accompanied by a PCR test; for emergency department visits this figure was 15%, and for outpatient visits this figure was only 3%.

Conclusion

We identified variations in the validity of diagnoses of SARS-CoV-2 infection recorded in different health care settings, geographic areas and over time. The performance of diagnosis codes for COVID-19 case ascertainment was better for inpatient and emergency department than for outpatient administrative data in the first year of the pandemic, likely owing to greater standardization in diagnosis coding practices in the former. However, over this period and subsequently, there were changes in guidelines and testing practices that may influence generalizability. Nevertheless, this study provides valuable insights about the validity of administrative data sources for COVID-19 case ascertainment that can benefit population-based research and surveillance.

References

1. Kluberger SA, Hou L, Dutcher SK, et al. Validation of diagnosis codes to identify hospitalized COVID-19 patients in health care claims data. *Pharmacoepidemiol Drug Saf* 2022;31:476-80.
2. Kluberger S. Validation of claims-based algorithms to identify hospitalized COVID-19 events within the FDA Sentinel System. ICPE all access special COVID-19 sessions; 2020 Nov. 6–Dec. 3; virtual meeting.
3. Wu G, D'Souza AG, Quan H, et al. Validity of ICD-10 codes for COVID-19 patients with hospital admissions or ED visits in Canada: a retrospective cohort study. *BMJ Open* 2022;12:e057838.
4. Kadri SS, Gundrum J, Warner S, et al. Uptake and accuracy of the diagnosis code for COVID-19 among US hospitalizations. *JAMA* 2020;324:2553-4.
5. Bhatt AS, McElrath EE, Claggett BL, et al. Accuracy of ICD-10 diagnostic codes to identify COVID-19 among hospitalized patients. *J Gen Intern Med* 2021;36:2532-5.
6. Hinds A, Lix LM, Smith M, et al. Quality of administrative health databases in Canada: a scoping review. *Can J Public Health* 2016;107:e56-61.
7. Lix LM, Vasylyk V, Ayilara O, et al. A synthesis of algorithms for multi-jurisdiction research in Canada. *Int J Popul Data Sci* 2022;7:1911.
8. ICD-10-CA coding direction for suspected COVID-19 cases. Ottawa: Canadian Institute for Health Information; 2020. Available: <https://www.cihi.ca/sites/default/files/document/covid-19-presentation-en.pdf> (accessed 2022 June 16).
9. Re: COVID-19 Temporary Fee Schedule Codes Implemented-Physicians can begin to submit claims for COVID-19 on May 1, 2020 [INFOBulletin]. Toronto: Health Services Branch, Ministry of Health; modified 2020 May 1. Available: <https://www.ontario.ca/document/ohip-infobulletins-2020/bulletin-4755-covid-19-temporary-fee-schedule-codes-implemented> (accessed 2022 June 1).
10. Virtual Telehealth Reference Guide billing. Vancouver: Divisions of Family Practice; 2020. Available: https://divisionsbc.ca/sites/default/files/54972/Virtual%20Telehealth%20Reference%20Guide%20-%20billing_0.pdf (accessed 2022 June 16).

11. von Elm E, Altman DG, Egger M, et al.; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:e296.
12. Benchimol EI, Smeeth L, Guttmann A, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
13. Lix LM, Walker R, Quan H, et al.; CHEP-ORTF Hypertension Outcomes and Surveillance Team. Features of physician services databases in Canada. *Chronic Dis Inj Can* 2012;32:186-93.
14. Claims processing solution (CPS). Winnipeg: Government of Manitoba. Available: <https://www.gov.mb.ca/health/claims/index.html> (accessed 2022 June 16).
15. Buajitti E, Chiodo S, Rosella LC. Agreement between area- and individual-level income measures in a population-based cohort: implications for population health research. *SSM Popul Health* 2020;10:100553.
16. Roos NP, Mustard CA. Variation in health and health care use by socioeconomic status in Winnipeg, Canada: Does the system work well? Yes and no. *Milbank Q* 1997;75:89-111.
17. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5.
18. Hale T, Angrist N, Goldsmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nat Hum Behav* 2021;5:529-38.

Affiliations: Department of Community Health Sciences (Lix), University of Manitoba, Winnipeg, Man.; Lady Davis Institute (Renoux, Moriello); Departments of Neurology and Neurosurgery, and Epidemiology, Biostatistics and Occupational Health (Renoux), McGill University, Montréal, Que.; George & Fay Yee Centre for Healthcare Innovation (Choi), University of Manitoba, Winnipeg, Man.; Department of Anesthesiology, Pharmacology & Therapeutics (Dormuth, Fisher), University of British Columbia, Vancouver, BC; Manitoba Centre for Health Policy (Dahl), University of Manitoba, Winnipeg, Man.; ICES Central (Wu, Asaf, Paterson), Toronto, Ont.

The Canadian Network for Observational Drug Effect Studies (CNODES) Investigators: Samy Suissa (principal investigator); Colin R. Dormuth (British Columbia); Brenda R. Hemmelgarn (Alberta); Jacqueline Quail (Saskatchewan); Dan Chateau (Manitoba); J. Michael Paterson (Ontario); Jacques LeLorier (Québec); Adrian R. Levy (Atlantic: Nova Scotia, Newfoundland and Labrador, New Brunswick, Prince Edward Island); Pierre Ernst and Kristian B. Filion (UK Clinical Practice Research Datalink); Lisa Lix (database); Robert W. Platt (methods); and Ingrid S. Sketris (knowledge translation).

Contributors: Lisa Lix, Christel Renoux, Carolina Moriello, Colin Dormuth, Anat Fisher, Matthew Dahl, Fangyun Wu, Ayesha Asaf and J. Michael Paterson were involved in data extraction, study design, data analysis and interpretation of results, and critically reviewed the manuscript for important intellectual content. Ko Choi was involved in data analysis and interpretation of results, and critically reviewed the manuscript for important intellectual content. Lisa Lix, Carolina Moriello, J. Michael Paterson and Ko Choi drafted the manuscript. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Funding: The Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating centre of the Drug Safety and Effectiveness Network, is funded by the Canadian Institutes of Health Research (CIHR; grant no. DSE-146021). The funders had no role in the design of the study, analysis and interpretation of data, and in writing of the manuscript.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is non-commercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Data sharing: This study was conducted by CNODES using administrative health data obtained through data-sharing agreements between its member research centres and their respective provincial data stewards. Data availability thus differs by site. British Columbia: The authors do not have permission to share data from this study. The data that support the findings of this study are available from Population Data BC (<https://www.popdata.bc.ca/>), but restrictions apply to the availability of these data, which were used under licence for the current study and so are not publicly available. Manitoba: Data used in this article were derived from administrative health and social data as a secondary source. The data were provided under specific data-sharing agreements only for approved use at Manitoba Centre for Health Policy (MCHP). The original source data are not owned by the researchers or MCHP and as such cannot be provided to a public repository. The original data source and approval for use have been noted in the acknowledgments of the article. Where necessary, source data specific to this article or project may be reviewed at MCHP with the consent of the original data providers, along with the required privacy and ethical review bodies. Ontario: The data set from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., health organizations and government) prohibit ICES from making the data set publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at <https://www.ices.on.ca/DAS> (email: das@ices.on.ca).

Acknowledgements: This study was made possible through data-sharing agreements between the CNODES member research centres and the respective provincial governments of British Columbia, Manitoba (Health Information Privacy Committee [HIPIC] no. 2021-378), and Ontario.

Disclaimer: The BC Ministry of Health approved access to and use of BC data facilitated by Population Data BC, for this study. All inferences, opinions and conclusions drawn in this manuscript are those of the authors, and do not reflect the opinions or policies of the data stewards. British Columbia data sources were as follows (<https://www2.gov.bc.ca/gov/content/health/conducting-health-research-evaluation/data-access-health-data-central>): British Columbia Ministry of Health [creator] (2022): Medical Services Plan (MSP) Payment Information File. BC Ministry of Health [publisher]. BC Ministry of Health (2022); Canadian Institute for Health Information [creator] (2022): National Ambulatory Care Reporting System. BC Ministry of Health [publisher]. BC Ministry of Health (2022); Canadian Institute for Health Information [creator] (2022): Discharge Abstract Database (Hospital Separations). BC Ministry of Health [publisher]. BC Ministry of Health (2022); British Columbia Ministry of Health [creator] (2022): Consolidation File (MSP Registration & Premium Billing). BC Ministry of Health [publisher]. BC Ministry of Health (2022). Parts of this material are based on data and/or information compiled and provided by the Canadian Institute for Health Information (CIHI). This study was supported by ICES, which is funded in part by an annual grant from the Ontario Ministry of Health. The authors also acknowledge the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under project no. 2022-004 (HIPIC no. 2021/2022-25). Data used in this study are from the Manitoba Population Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided by Manitoba Health. The opinions, results and conclusions are those of the authors. No endorsement by the provincial governments, data stewards, CIHI, Health Canada or CIHR is intended or should be inferred.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/11/5/E790/suppl/DC1.