BMJ Open

Identifying people with post-COVID condition using linked, population-based administrative health data from Manitoba, Canada: Prevalence and predictors in the COVID-positive population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-087920
Article Type:	Original research
Date Submitted by the Author:	22-Apr-2024
Complete List of Authors:	Katz, Alan; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences; Rady Faculty of Health Sciences University of Manitoba, Department of Family Medicine Ekuma, Okechukwu; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Enns, Jennifer; University of Manitoba Faculty of Health Sciences, Manitoba Centre for Health Policy Cavett, Teresa; University of Manitoba Singer, Alexander; Rady Faculty of Health Sciences, University of Manitoba, Family Medicine Sanchez-Ramirez, Diana C.; University of Manitoba; University of Manitoba Keynan, Yoav; University of Manitoba, Medical Microbiology and Infectious Diseases Lix, Lisa; University of Manitoba, Community Health Sciences Walld, Randy; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Yogendran, Marina; Manitoba Centre for Health Policy Nickel, Nathan C.; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Urquia, Marcelo; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Star, Leona; First Nations Health and Social Secretariat of Manitoba Olafson, Kendiss; Rady Faculty of Health Sciences University of Manitoba, Department of Internal Medicine Logsetty, Sarvesh; University of Manitoba, Departments of Surgery and Psychiatry Spiwak, Rae; University of Manitoba, Department of Surgery Waruk, Jillian; First Nations Health and Social Secretariat of Manitoba Matharaarachichi, Surani; University of Manitoba, Department of Statistics
Keywords:	SARS-CoV-2 Infection, COVID-19, Post-Acute COVID-19 Syndrome, Epidemiology < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

Title:

Identifying people with post-COVID condition using linked, population-based administrative health data from Manitoba, Canada: Prevalence and predictors in the COVID-positive population

Authors and Affiliations:

Alan Katz[1,2], Okechukwu Ekuma[1], Jennifer E. Enns[1], Teresa Cavett[2], Alexander Singer[2], Diana C. Sanchez-Ramirez[3], Yoav Keynan[4], Lisa M. Lix[1], Randy Walld[1], Marina S. Yogendran[1], Nathan Nickel[1], Marcelo L. Urquia[1], Leona Star[5], Kendiss Olafson[4], Sarvesh Logsetty[6], Rae Spiwak[6], Jillian Waruk[5], Surani Matharaarachichi[7]

1. Manitoba Centre for Health Policy, Department of Community Health Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
2. Department of Family Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
3. College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
4. Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
5. First Nations Health and Social Secretariat of Manitoba, Winnipeg, Canada
6. Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
7. Department of Statistics, University of Manitoba, Winnipeg, Canada

Corresponding Author:

Dr. Alan Katz
Manitoba Centre for Health Policy
University of Manitoba
408-727 McDermot Ave
Winnipeg, Manitoba
R3E 3P5
Alan.Katz@umanitoba.ca

Abstract

Objective: Many individuals exposed to SARS-CoV-2 experience long-term symptoms as part of a syndrome called post-COVID condition (PCC). Research on PCC is still emerging but is urgently needed to support diagnosis, clinical treatment guidelines, and health system resource allocation. In this study, we developed a method to identify PCC cases using administrative health data, and report PCC prevalence and predictive factors in Manitoba, Canada.

Design: Cohort study.

Setting: Manitoba, Canada.

Participants: All Manitobans who tested positive for SARS-CoV-2 during population-wide PCR testing from March 2020 to December 2021 (n=66,365), and were subsequently deemed to have PCC based on ICD-9/10 diagnostic codes and prescription drug codes (n=11,316). Additional PCC cases were identified using predictive modelling to assess patterns of health service use, including physician visits, emergency department visits, and hospitalizations (n=4,155).

Outcomes: We measured PCC prevalence (% PCC cases among Manitobans with positive tests) and identified predictive factors associated with PCC (odds ratios with 95% confidence intervals, adjusted for socio-demographic and clinical characteristics).

Results: Among 66,365 Manitobans with positive tests, we identified 15,471 (23%) as having PCC. Being female (aOR 1.64, 95% CI 1.58-1.71), age 59 or older (age 60-79 aOR 1.33, 95% CI 1.25-1.41; age 80+ aOR 1.62, 95% CI 1.46-1.80), hospitalized within 14 days of COVID-19 infection (aOR 1.95, 95% CI 1.80-2.10), and having a Charlson Comorbidity Index of 1+ (aOR 1.95, 95% CI 1.78-2.14) were predictive of PCC. Receiving 1+ doses of the COVID-19 vaccine (1 dose aOR 0.80, 95% CI 0.74-0.86; 2 doses aOR 0.29, 95% CI 0.22-0.31) decreased the odds of PCC.

Conclusions: This data-driven approach expands our understanding of the prevalence and epidemiology of PCC and may be applied in other jurisdictions with population-based data. The study provides additional insights into risk and protective factors for PCC to inform health system planning and service delivery.

Keywords: SARS-CoV-2; post-COVID condition; long COVID; administrative health data; data linkage; epidemiology; predictive modeling; Canada

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

Funding

The study received funding from a Canadian Institutes of Health Research Operating Grant: Emerging COVID-19 Research Gaps and Priorities (Grant No. GA7-1777787) awarded to Alan Katz in 2020. The results and conclusions are those of the authors and no official endorsement by the funder is intended or should be inferred.

Competing Interests

The authors declare that they have no competing interests.

Strengths and Limitations

- The Manitoba Population Research Data Repository holds an extensive collection of administrative datasets that are population-based and linkable across domains at the individual and family levels, making it a valuable resource for population health research.
- We had access to multiple sources of relevant data (COVID-19 test and vaccination data, physician billing claim codes, prescribed medications, and physician visit and hospitalization patterns) from which to triangulate our approach and identify people with post-COVID condition.
- Despite the strengths of this approach, we cannot be certain that all of the symptoms and health service use patterns we identified as post-COVID condition were all directly related, since providers and patients may not have recognized their symptoms as being post-COVID condition at the time of the visit, or may have been seeking medical help for symptoms that were not actually post-COVID condition.
- Using natural language processing to incorporate free text from physicians' clinical notes into our analysis would provide additional context and nuance to our administrative data-driven approach, and may be an option for future research in this area.

Introduction

The SARS-CoV-2 pandemic created significant societal disruption across the globe, with over 774 million confirmed cases and more than 7 million deaths as of early 2024.[1] While COVID-19 is primarily considered an acute illness, some individuals experience long-term symptoms in a syndrome referred to as 'long COVID' or 'post-COVID condition' (PCC). More than 200 symptoms have been reported to occur in the weeks following SARS-CoV-2 infection, including shortness of breath, cognitive dysfunction, and fatigue, as well as multiple conditions that impact daily functions, such as neurological and psychiatric disorders, hypertension, respiratory difficulties, heart palpitations, chest pain, myalgia, and digestive disorders.[2–4]

Our current understanding of PCC is derived from a series of clinical cohort, cross-sectional, and survey-based studies.[5–11] Estimates of PCC prevalence vary, with 32-87% of COVID-19 survivors reporting persistent or new symptoms two to three months after acute infection.[12] Many individuals with PCC are young, previously healthy, and had mild forms of COVID-19 that did not require hospitalization.[13] There appears to be higher prevalence of PCC in particular groups, including females and individuals aged 35-45.[13–15] SARS-CoV-2 infection has also been shown to disproportionately impact many communities who may be lower income, have higher rates of comorbidities, live in crowded or inadequate housing conditions, and have poorer access to health services than the general population.[16] This could mean a higher prevalence of PCC in, for example, Indigenous populations or immigrant groups.[17,18]

PCC can have a tremendous impact on quality of life and has been shown to substantially increase health service use.[19–21] The underlying mechanism of PCC is now presumed to be mediated by the immune system.[22,23] The link between complement

dysregulation and PCC explains the three principal mechanisms: immune dysregulation, immune priming, and microvascular blood clotting[23]; as well, the activation of cytomegalovirus- and Epstein-Barr virus-specific antibodies may also contribute to the pathophysiology of PCC.[23,24]

Despite PCC becoming better understood, several important clinical questions remain. Due to the emerging nature of the illness, the large number of associated symptoms, and the diverse methodologies used thus far to study PCC, there is currently no consensus on the true clinical definition of PCC. While the World Health Organization (WHO), the Government of Canada, the National Institute for Clinical Excellence, and the Centres for Disease Control and Prevention have each proposed different clinical definitions, these may continue to undergo refinement as the understanding of PCC and its pathophysiology evolves. In the meantime, an analytic approach for identifying PCC in the general population would be a significant asset, since it would support the investigation of risk characteristics, medium- to long-term health outcomes, and recovery trajectories of those with PCC. This evidence is critical for informing the development of clinical treatment guidelines and health system resource planning and would also aid in the development of COVID-related health policy worldwide.

Several editorials have called for data science-based solutions to address some of the challenges the pandemic has presented.[25–27] Administrative data are a powerful resource for researchers, clinicians, patients, and health system decision-makers, and are currently being used to help us understand COVID-19 and its impact across various populations. An administrative data approach has the advantage of capturing a whole population of interest, thereby limiting selection bias and loss to follow-up, and because administrative data include community-based care, they allow for the inclusion of sub-populations with less severe

disease that may not be present in hospital or critical care studies. In this study, we constructed a PCC cohort using administrative health data from a population-based data repository in Manitoba, Canada, and described PCC risk and protective factors in the COVID-positive population.

Methods

Study Setting

Manitoba is a province with 1.4 million residents in the geographical centre of Canada. The provincial single-payer healthcare system covers over 99% of Manitoba residents, excluding only individuals who are federally insured (e.g., those incarcerated in federal prisons, members of the military, and some First Nations and Inuit populations).[28,29] The provincial health insurance coverage includes all hospitalizations, medically necessary physician services, and prescription drug dispensations by Manitoba pharmacies.

Data Sources

The Manitoba Population Research Data Repository at the Manitoba Centre for Health Policy comprises over 100 databases of whole-population, individual-level administrative data from the health, social services, legal, and education systems in Manitoba. All Repository databases are de-identified (names and addresses removed), but they are linkable at the individual and family level using a scrambled Patient Health Identification Number attached to each record and to a central population registry.[29,30] The Repository data have been used extensively in research and the validity of the databases to examine population health has been well documented.[31–33] The databases used in this study are presented in **Table 1**. Both the hospital discharge abstract data and the medical claims data have been shown to be comprehensive for the population of Manitoba.

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

Table 1. Databases and key variables from the Manitoba Population Research Data Repository

Database	Key Variables or Information
Manitoba Health Insurance Registry	Biological sex, date of birth, and postal code for the population of Manitoba
COVID-19 Laboratory Testing and Results	COVID-19 PCR test information (date, test result)
COVID-19 Vaccinations	COVID-19 vaccinations (date, type of vaccine, number of doses)
Hospital Discharge Abstracts	ICD-10 diagnosis and procedure codes from in-patient hospital stays
Hospital Admission, Discharge, and Transfer Database	Data on hospital and ICU admission, discharge, and transfer
Medical Claims/Medical Services	Billing claims from primary care clinics (physicians and nurse practitioners) and specialist physicians, including ICD-9 diagnosis and procedure codes
Drug Program Information Network	All prescription medications dispensed by community retail pharmacies
Emergency Department Information System	Data on emergency department visits
Immigration, Refugees, and Citizenship Canada (IRCC) Permanent Resident Database	Data on people who immigrated to Manitoba
Manitoba First Nations Research File	People registered as First Nations according to the Indian Act of Canada

ICD: International Classification of Disease; ICU: intensive care unit; PCR: Polymerase Chain Reaction

Patient/Public Involvement

At least one member of the research team was an individual who was experiencing PCC at the time of the study. The study conception, design, conduct, and interpretation were informed by their personal experiences with PCC and as a family physician. They are also listed as an author on the study.

Ethics and Privacy

Ethics approvals were granted by the University of Manitoba's Health Research Ethics Board (#HS25090 H2021:279) and the Health Information Research Governance Committee at the Manitoba First Nations Health and Social Secretariat (2020). The study protocol was reviewed by the Manitoba Government Health Information Privacy Committee (HIPC #2021/2022–18).

Constructing the PCC Cohort

The cohort construction comprised three key steps, an overview of which is presented in **Figure 1**. All cohort construction steps and analyses were completed in SAS/STAT® software version 15.1.[34]

Step 1. Identify the population at risk for PCC.

We began by identifying all Manitobans at risk for developing PCC, i.e. those with a positive COVID-19 test result from March 1, 2020 to December 31, 2021 (Step 1 cohort). The province of Manitoba provided whole-population COVID-19 PCR testing from March 2020 to December 2021, thus we were able to include information on the *alpha*, *delta,* and early *omicron* waves of infection. The province's centralized testing approach was discontinued on December 31, 2021.

Step 2. Use diagnosis codes and prescription drug codes to narrow to those with PCC.

From the rapidly evolving literature on COVID-19,[3,20,35–37] we gathered information on symptoms commonly associated with PCC and medications used to treat these symptoms. We determined the corresponding International Classification of Disease (ICD)-9 (out-patient) and ICD-10 (in-hospital) diagnosis and procedure codes and Anatomical Therapeutic Chemical (ATC) codes (**Table S1**), and used these to narrow the cohort to only people who reported new symptoms or received a new prescription related to PCC symptoms 90 days or more after their positive COVID-19 test date (Step 2 cohort). We excluded anyone who reported the same symptoms in the three years before their positive COVID-19 test. However, given the broad range of symptoms associated with COVID-19, we knew this step likely wouldn't be exhaustive, because individuals with PCC might not have reported new symptoms (many are common to other illnesses) and might not have received a new prescription if their symptoms were ones they had experienced previously. Thus, we used

another strategy to identify additional individuals in the Step 1 cohort likely to have experienced PCC but not captured in Step 2.

Step 3. Identify additional individuals with PCC based on predictive modeling.

In this final step, we used a predictive modeling approach to identify individuals likely to have PCC based on their physician visit rates. First, in the PCC cohort from Step 2, we calculated the physician visit rate from the 91st day after their first positive COVID-19 test date to March 31, 2022 (PCC period), and compared it to their visit rate from April 1, 2019 to March 31, 2020 (pre-COVID period). We also compared physician visit rates for these two periods among those that tested positive for COVID-19 in Step 1 but were not part of the Step 2 cohort. In this latter group, those with visit rates the same or greater than those of the Step 2 cohort were added to the final cohort. Then, we trained several different regression model builds (step-wise, best subset, and lasso) to determine the best fit. Model selection was done by considering the Brier score, the Schwarz Bayesian Criterion, and the C-statistic,[38,39] and the selected model was then internally validated by bootstrapping 500 samples, yielding an optimism-corrected C-statistic of 0.69387.[40] This method ensured that we used all the data we had available in the prediction process. The selected model was then re-fitted to the entire group of individuals who tested positive for COVID-19 (Step 1 cohort) to determine the predicted probability of any of those individuals having PCC. The list of variables considered in these models and the ones that were used in the selected model are shown in **Table S2.**

One of the notable challenges in predictive modeling is determining the optimal cut-off with which to classify individuals as predicted to have the outcome. We tested several methods to obtain the highest optimal cut-point for a more conservative classification,[41] and then classified the Step 1 cohort into those predicted to have PCC and those not predicted to have PCC.

Validating Step 3 of Cohort Construction

Our efforts to identify individuals with PCC in Step 3 were validated in a three-way comparison: we compared three health service use indicators (physician visit rates, hospitalizations, and emergency department visits) in the Step 2 cohort, the additional individuals identified in Step 3, and a group matched on a High Dimensional Propensity Score (HDPS) that encompassed both socio-demographic characteristics and health service use data dimensions. We hypothesized that the Step 3 cohort would have health service use patterns similar to the Step 2 cohort.

High Dimensional Propensity Score Matching

Following the method developed by Li et al.,[42] we hard-matched in a ratio of 25:1 on age, sex, and region of residence (urban/rural) at the time of first positive COVID-19 test and calculated standardized differences on these variables. Then, we generated HDPS using demographic variables (age, sex, region of residence, income quintile) and health service use data dimensions (hospital diagnosis, hospital procedures, physician diagnosis, physician tariff codes, ATC codes for prescription drugs) (**Figure S1**). The HDPS models were stratified by sex and run separately for three health service use indicators: physician visits, hospitalizations, and emergency department visits. To create the final matched group, the five closest matches to the HDPS in each pool of 25 were selected. A caliper of 0.2× the pooled standard deviation of the logit (p-score) was applied to avoid selecting pairs with unsuitably large differences in p-score.

Statistical Analysis

The results of the three-way comparison used to validate the Step 3 cohort creation process are shown in **Figure 2**. We examined physician visit rates using negative binomial distribution with the log of person-days as the offset, obtaining rate ratios and 95% confidence intervals, and we conducted time-to-event analyses of first hospitalizations and first emergency department visits after index date, obtaining hazard ratios and 95% confidence intervals. The time-to-event approach was used for the latter two indicators because they were rare events.

Predictors of Post-COVID Condition: Univariate and Multivariate Analyses

We used univariate and multivariate logistic regression modeling to identify socio-demographic and clinical variables that were predictive or protective of PCC, producing odds ratios and 95% confidence intervals. Each of the univariate models had PCC as the dependent variable and one of the following as the independent variable: sex (male/female); age group (0-18, 19-29, 30-59, 60-79, 80+); region of residence (urban/rural); income quintile (Q1-Q5); Regional Health Authority (Winnipeg, Interlake-Eastern Regional Health Authority, Prairie Mountain Health, Southern Health/Santé-Sud, Northern Health Region); hospitalization within 14 days of positive COVID-19 test (Y/N); number of COVID-19 vaccine doses before first positive COVID-19 test (0-3+); comorbidity (Charlson Comorbidity Index, 0-3+); immigrant to Manitoba within the last five years (Y/N); and First Nations (Y/N). The multivariate models included all of the variables in the list above.

Results

PCC Prevalence

As shown in **Figure 1**, we identified 66,365 individuals who tested positive for COVID-19 from March 1, 2020 to December 31, 2021 in Step 1. Among these individuals, 11,316 (17.1%) were identified as having PCC based on their diagnostic codes and prescription drug codes in Step 2. Physician visit patterns for the remaining 55,049 individuals who tested positive for COVID-19 were then further examined in Step 3. The pre-COVID vs PCC period physician visit rate ratio for the Step 2 cohort (N=11,316) was 1.75; among the other 55,049 individuals who tested positive for COVID-19, we identified 4,155 (7.5% of 55,049) individuals as likely having PCC in Step 3 based on their physician visit rate and/or the predictive modeling. This group was added to the Step 2 cohort for a total of 15,471 PCC cases. Thus, we determined the prevalence of PCC to be 23.3% in the COVID-positive population of Manitoba.

Rate ratios for physician visits and hazard ratios for hospitalizations and emergency department visits were calculated as part of the validation process for constructing the final PCC cohort to determine whether there were significant differences in health service use at each step. Shown in **Figure 2**, these data confirm that the rates of health service use were either the same or slightly higher between the Step 2 and Step 3 cohorts.

Predictors of Post-COVID Condition

In the final PCC cohort, we conducted univariate and multivariate analyses to determine the odds of developing PCC based on the cohort's sociodemographic characteristics. The univariate analyses are presented in **Table 2** as odds ratios and 95% confidence intervals**.** Female individuals and those over the age of 59 had higher odds of

developing PCC, as did people in lower income brackets, people with comorbidities (measured by the Charlson Comorbidity Index), and people who were hospitalized within 14 days of a positive COVID-19 test. First Nations also had higher odds of developing PCC than other Manitobans, whereas immigrants who arrived in Manitoba within the last five years had lower odds. The COVID-19 vaccine was a protective factor and a higher number of vaccine doses (from 0-3+) conferred greater protection against PCC.

The multivariate analyses are shown in **Table 3**. Even after adjusting for age, sex, income, and region of residence, many of the same factors were predictive of developing PCC, but a few discrepancies between the univariate and the multivariate findings are notable: in the multivariate findings, lower income was no longer a significant predictor of PCC, nor was living in the more remote regions of the province (e.g., the Northern Health Region).

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

Table 2. Predictive characteristics for developing PCC in the population of Manitoba, Canada

Univariate odds ratios and 95% confidence intervals

	Odds Ratio	95% CI
Sex		
Male (ref)		
Female	1.60	1.54, 1.66
Age group (years)		
0-18	0.40	0.38, 0.42
19-29	0.87	0.83, 0.91
30-59 (ref)		
60-79	1.54	1.46,1.63
80+	2.20	2.01, 2.41
Region of residence		
Urban (ref)		
Rural	0.92	0.89, 0.96
Income quintile		
Q1 (lowest)	1.32	1.24, 1.40
Q2	1.24	1.17, 1.32
Q3	1.14	1.07, 1.22
Q4	1.10	1.03, 1.17
Q5 (highest) (ref)		
Not found	2.13	1.78, 2.56
Missing	1.72	1.43, 2.07
Regional Health Authority		
Winnipeg Regional Health Authority (ref)		
Interlake-Eastern Regional Health Authority	0.86	0.80, 0.92
Northern Health Region	1.11	1.05, 1.18
Southern Health-Santé Sud	0.80	0.76, 0.84
Prairie Mountain Health	0.98	0.91, 1.04
Charlson Comorbidity Index		
0 (no comorbidity) (ref)		
1	1.79	1.71, 1.88
2	2.14	1.98, 2.33
3+ (higher comorbidity)	3.20	2.94, 3.47
Hospitalization within 14 days of positive COVID-19 test	3.28	3.05, 3.52
Number of COVID-19 vaccine doses before positive COVID-19 test		
0 doses (ref)		
1 dose	0.93	0.87, 0.99
2 doses	0.34	0.32, 0.36
3+ doses	0.35	0.30, 0.40
Recently immigrated to Manitoba (within 5 years of the study period)	0.84	0.77, 0.91

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

First Nations	1.20	1.15, 1.25

Table 3. Predictive characteristics for developing PCC in the population of Manitoba, Canada

Multivariate odds ratios and 95% confidence intervals

	Odds Ratio	95% CI
Sex		
Male (ref)		
Female	1.64	1.58, 1.71
Age group (years)		
0-18	0.37	0.35, 0.39
19-29	0.92	0.87, 0.97
30-59 (ref)		
60-79	1.33	1.25, 1.41
80+	1.62	1.46, 1.80
Region of residence		
Urban (ref)		
Rural	0.93	0.81, 1.06
Income quintile		
Q1 (lowest)	1.00	0.94, 1.07
Q2	0.99	0.93, 1.06
Q3	1.01	0.95, 1.08
Q4	0.99	0.92, 1.06
Q5 (highest) (ref)		
Not found	0.62	0.51, 0.76
Unknown	1.15	0.94, 1.40
Regional Health Authority		
Winnipeg Regional Health Authority (ref)		
Interlake-Eastern Regional Health Authority	0.90	0.78, 1.05
Northern Health Region	1.12	0.97, 1.30
Southern Health-Santé Sud	0.80	0.69, 0.92
Prairie Mountain Health	1.11	1.00, 1.23
Charlson Comorbidity Index		
0 (no comorbidity) (ref)		
1	1.40	1.32, 1.47
2	1.46	1.33, 1.60
3+ (higher comorbidity)	1.95	1.78, 2.14
Hospitalization within 14 days of positive COVID-19 test	1.95	1.80, 2.10
Number of COVID-19 vaccine doses before positive COVID-19 test		
0 (ref)		
1	0.80	0.74, 0.86
2	0.29	0.22, 0.31
3+	0.26	0.22, 0.30

/ScholarOne/conversions/709678896889306151 7/54507837_File000007_1334835312.docx

Recently immigrated to Manitoba (within 5 years of the study period)	0.88	0.80, 0.96
First Nations	1.12	1.06, 1.18

Discussion

Our study aimed to develop an administrative data-driven analytic approach to identifying PCC cases in the general population. We identified PCC cases by first identifying individuals with a positive COVID-19 PCR test, then used diagnostic codes, prescription drug codes, and patterns of health service use to develop a cohort of 15,471 individuals with PCC, which represents a prevalence of 23.3% in the segment of the Manitoba population who tested positive for COVID-19. In this cohort, individuals who were female, age 59+, First Nations, had one or more comorbidity, or were hospitalized within two weeks of a positive COVID-19 test had higher odds of developing PCC than other Manitobans. Vaccination for COVID-19 (one or more doses) was protective against PCC.

In the literature, we observed that others using administrative data to examine PCC prevalence generally sought to define PCC using a combination of health provider claims, hospital discharge abstracts, and/or electronic health records.[43,44] A study from British Columbia, Canada, used data from patients attending a PCC-specific clinic and patients with an ICD-10-coded hospital admission for PCC, and then identified other potential PCC cases with similar demographics, pre-existing conditions, and COVID-19 symptoms using an elastic net penalized logistic regression model; they reported a PCC prevalence of 18% in the COVID-positive population.[45] The authors acknowledged that they likely underestimated the prevalence of PCC since their analyses were based on the characteristics of severely impacted patients who were either hospitalized or referred to the specialized PCC clinic. However, they had the advantage of established ICD-10 codes, whereas we did not have a specific code for PCC at the time of our study. Other possible reasons for the difference in estimates between ours and others' are the lag time between a positive PCR test and the

development of PCC symptoms (some authors used 30-60 days,[46,47] while we used the WHO definition of 90 days[48]) and the severity of illness in the starting population (e.g., PCC among people hospitalized for acute COVID-19[49]).

The major challenge we encountered in developing our analytic approach was that people experiencing PCC symptoms were not readily identified in the Manitoba Population Research Data Repository. Without access to a specific ICD code for PCC, we instead used surrogate markers (physician billing claim codes, prescribed medications, and physician visit and hospitalization patterns), but we cannot be certain that these were all directly related to PCC. In addition, outpatient billing practices in Manitoba capture only one main diagnosis, potentially limiting the information we had available to identify PCC cases. As well, a possible lack of awareness or understanding of PCC at the time of providing the clinical care and completing the billing claims could have resulted in providers not screening for persistent symptoms and patients not seeking medical help specifically for PCC, especially if symptoms were mild. In an online survey of Manitobans in 2022, 62.6% indicated that PCC symptoms impacted their daily lives, but only 50.8% sought healthcare (among these, 65.9% primary care, 15.2% emergency department, and 32.0% specialist or therapist).[50] To account for these limitations, we relied on the breadth of data available in the repository, drawing from multiple linked databases and using analytic techniques to assemble the PCC cohort. At each analytic step, the research team discussed how the results aligned with their clinical observations and with the literature, and this discussion informed the next step.

A second challenge in our approach was that we weren't certain which variables should be included in the predictive models to identify potential PCC cases, and it's possible that our approach overestimated the prevalence of PCC by including false positives in the cohort. As more research on PCC becomes available, our model could be adjusted to include

additional variables, thus increasing sensitivity, and leading to a more refined classification of PCC cases. It would also likely be helpful to include clinical notes in our administrative data definition of PCC. Our structured administrative claims data currently lack the context that would be provided by linked clinical notes from physician visits. Recent advances in natural language processing would facilitate the incorporation of free text into analyses of structured administrative data.[51–53]

With respect to PCC predictors, we identified several sociodemographic factors that were associated with higher odds of developing PCC. Several other studies have reported similar findings, particularly the higher odds of PCC among female and/or older individuals, people who were severely ill after contracting acute COVID-19 (e.g., required hospitalization), and people with certain types of comorbidities.[52,54–58] Higher odds of PCC among First Nations has not been reported previously, although there is ample published research showing how colonial practices and policies have disadvantaged First Nations and other Indigenous populations over many generations and placed them at risk for poor health outcomes, particularly in a public health crisis like the SARS-CoV-2 pandemic.[59–61] A recent study reports on the public health measures implemented in Manitoba, how First Nations acted to protect their communities from COVID-19, and their success in advocating for vaccine prioritization.[62] Immigrants having lower odds of developing PCC than other Manitobans can perhaps be explained by the healthy immigrant effect,[18] although examining sub-groups of immigrants would likely provide more insight into this finding. In our current study, vaccination was found to be protective against PCC, and this finding is generally supported by existing literature,[52,55,57,63] although a 2023 study by Luo et al[54] did not find a significant decrease in PCC risk with more vaccine doses, possibly because the study

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

examined a later time period than ours, where most individuals already had 3+ doses of the COVID-19 vaccine.

The method we present here for estimating the prevalence of PCC using linked administrative data is widely generalizable to other jurisdictions with access to health system data, and will serve as the basis for further research (ours and others') to follow the natural history and health outcomes of PCC. In particular, we plan to examine PCC outcomes in Indigenous and immigrant populations in more detail. Our findings are being shared widely within our research team's networks of healthcare providers and planners to support the resourcing and implementation of health services for patients living with PCC. As well, the ongoing discourse on the pathogenetic mechanisms of PCC draws many parallels with other post-viral conditions,[64] and as such, our approach could inform future research to better understand the prevalence and course of these conditions. The constant genetic variation and the evolution of new variants and sub-variants and the changing application of testing and vaccination modalities limits our ability to describe the current prevalence of COVID-19 and PCC. However, as more data become available, we could conduct time series analyses based on the emergence of new variants to address this limitation.

Conclusion

Population-wide administrative data have the potential to address important gaps in our knowledge of emerging health conditions, and can provide nuanced information for clinical and health system recovery and prevention efforts in the aftermath of a global public health crisis like the SARS-CoV-2 pandemic. Their value may be further enhanced by the use of natural language processing to link unstructured clinical data to existing repositories, which would provide additional context to structured administrative data.

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

List of Abbreviations

ATC	Anatomical Therapeutic Chemical
HDPS	High Dimensional Propensity Score
ICD-9	International Classification of Disease (out-patient)
ICD-10	International Classification of Disease (in hospital)
PCC	Post-COVID Condition
PCR	Polymerase Chain Reaction
WHO	World Health Organization

Acknowledgements

The authors acknowledge the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under project #2021-033 (HIPC/PHRPC #2021/2022-18). 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 databases provided by Manitoba Health; Statistics Canada; Shared Health Manitoba; Immigration, Refugees, and Citizenship Canada; First Nations Health and Social Secretariat of Manitoba; and Manitoba Primary Care Research Network. The results and conclusions are those of the authors and no official endorsement by Manitoba Health or other data providers is intended or should be inferred.

We would also like to thank Annshirley Aba Afful, Quinn Anhorn, Dana Kliewer, Katie Kitchen, Marlee Mayer, Saman Muthukumaranra, and Gillian Martens for their contributions to analysis planning and manuscript writing.

Author Contributions

AK, JEE, TC, AS, DCS-R, YK, LML, NN, MLU, and LS were responsible for research conception; all of these authors as well as KO, SL, and RS contributed to study design. OE, RW, and MSY conducted the data analyses. All authors, including JW and SM, contributed to the interpretation of the analytic findings. JEE and AK drafted the manuscript. All authors reviewed the draft manuscript for intellectual content and gave approval for publication. All authors agree to be accountable for all aspects of the work.

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

Data Sharing Statement

The source data used in this study were originally collected during the routine administration of health and social services in Manitoba, and were provided to the Manitoba Centre for Health Policy (MCHP) for secondary use in research under specific data sharing agreements between the data trustees and MCHP. The data are approved for use at MCHP only. They are not owned by the researchers or by MCHP and cannot be deposited in a public repository. To review source data specific to this article or project, interested parties should contact the MCHP Repository Access & Use team at MCHP.Access@umanitoba.ca. The team will then facilitate data access by seeking the consent of the original data holders and the required privacy and ethics review bodies. Research studies using First Nations data require ethics approval from the Health Information Research Governance Committee at the First Nations Health and Social Secretariat of Manitoba (https://www.fnhssm.com/hirgc), and we comply with their policies for data access, linkage and sharing. For inquiries about accessing Manitoba First Nations data, please contact info@fnhssm.com.

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

References

1. World Health Organization. COVID-19 Epidemiological Update 2024. Coronavirus disease (COVID-19) weekly epidemiological updates and monthly operational updates. Published 2024. Accessed February 8, 2024. https://www.who.int/publications/m/item/covid-19-epidemiological-update---19-january-2024
2. Centers for Disease Control and Prevention. Post-COVID/Long COVID. Published 2020. Accessed March 29, 2021. https://www.idsociety.org/covid-19-real-time-learning-network/disease-manifestations--complications/post-covid-syndrome/
3. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine*. 2021;38:101019. doi:10.1016/J.ECLINM.2021.101019
4. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry*. 2021;8(5):416-427. doi:10.1016/S2215-0366(21)00084-5
5. Chen X, Norris C, Whitten T, Ho C, Mann B, Bakal J. Symptomology following COVID-19 among adults in Alberta, Canada: an observational survey study. *BMJ Open*. 2024;14(3):e078119. doi:10.1136/bmjopen-2023-078119
6. Cohen J, van der Meulen Rodgers Y. An intersectional analysis of long COVID prevalence. *Int J Equity Health*. 2023;22(1):261. doi:10.1186/s12939-023-02072-5
7. Egger M, Vogelgesang L, Reitelbach J, Bergmann J, Müller F, Jahn K. Severe Post-COVID-19 Condition after Mild Infection: Physical and Mental Health Eight Months Post Infection: A Cross-Sectional Study. *Int J Environ Res Public Health*. 2024;21(1). doi:10.3390/ijerph21010021
8. Hastie CE, Lowe DJ, McAuley A, et al. True prevalence of long-COVID in a nationwide, population cohort study. *Nat Commun*. 2023;14(1):7892. doi:10.1038/s41467-023-43661-w
9. Shah AD, Subramanian A, Lewis J, et al. Long COVID symptoms and diagnosis in primary care: A cohort study using structured and unstructured data in The Health Improvement Network primary care database. *PLoS One*. 2023;18(9 September):1-21. doi:10.1371/journal.pone.0290583
10. Tan S, Pryor AJG, Melville GW, Fischer O, Hewitt L, Davis KJ. The lingering symptoms of post-COVID-19 condition (long-COVID): a prospective cohort study. *Intern Med J*. 2024;54(2):224-233. doi:10.1111/imj.16251
11. Zhang Y, Chinchilli VM, Ssentongo P, Ba DM. Association of Long COVID with mental health disorders: a retrospective cohort study using real-world data from the USA. *BMJ Open*. 2024;14(2):1-8. doi:10.1136/bmjopen-2023-079267
12. Sugiyama A, Miwata K, Kitahara Y, et al. Long COVID occurrence in COVID-19 survivors. *Sci Rep*. 2022;12(1):6039. doi:10.1038/S41598-022-10051-Z
13. Rubin R. As their numbers grow, COVID-19 “long haulers” stump experts. *JAMA*. 2020;324(14):1381-1383. https://jamanetwork.com/
14. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med*. 2021;27(4):626-631. doi:10.1038/s41591-021-01292-y

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

15. Sudre CH, Murray B, Varsavsky T, et al. Author Correction: Attributes and predictors of long COVID. *Nature Medicine 2021 27:6*. 2021;27(6):1113-1120. doi:10.1038/s41591-021-01361-2

16. Siff EJ, Aghagoli G, Marin BG, Tobin-Tyler E, Poitevien P. SARS-CoV-2 transmission: a sociological review. *Epidemiol Infect*. 2020;148:e242. doi:10.1017/S095026882000240X

17. Power T, Wilson D, Best O, et al. COVID-19 and Indigenous Peoples: An imperative for action. *J Clin Nurs*. 2020;29(15-16):2737. doi:10.1111/JOCN.15320

18. Vang ZM, Ng E. The impacts of COVID-19 on immigrants and the healthy immigrant effect: Reflections from Canada. *Prev Med (Baltim)*. 2023;171:107501. doi:10.1016/J.YPMED.2023.107501

19. O' Mahony L, Buwalda T, Blair M, et al. Impact of Long COVID on health and quality of life. *HRB Open Res*. 2022;5. doi:10.12688/HRBOPENRES.13516.1

20. Mahbub Hossain M, Das J, Rahman F, et al. Living with "long COVID": A systematic review and meta-synthesis of qualitative evidence. *PLoS One*. 2023;18(2). doi:10.1371/JOURNAL.PONE.0281884

21. Wolf S, Zechmeister-Koss I, Erdös J. Possible long COVID healthcare pathways: a scoping review. *BMC Health Serv Res*. 2022;22(1):1-11. doi:10.1186/S12913-022-08384-6/FIGURES/2

22. Baillie K, Davies HE, Keat SBK, et al. Complement dysregulation is a predictive and therapeutically amenable feature of long COVID. *medRxiv*. Published online 2023:2023.10.26.23297597. https://www.medrxiv.org/content/10.1101/2023.10.26.23297597v1

23. Cervia-Hasler C, Brüningk SC, Hoch T, et al. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. *Science*. 2024;383(6680):eadg7942. doi:10.1126/science.adg7942

24. Klein J, Wood J, Jaycox JR, et al. Distinguishing features of long COVID identified through immune profiling. *Nature*. 2023;623(7985):139-148. doi:10.1038/s41586-023-06651-y

25. Kolaczyk ED, Lee MM, Liu J, Parker MS. We Need a (Responsible!) Data Science Rapid Response Network. *Harv Data Sci Rev*. 2021;3(3.1). doi:10.1162/99608f92.2794e78d

26. Leonelli S. Data Science in Times of Pan(dem)ic. *Harv Data Sci Rev*. 2021;3(3.1):2021. doi:10.1162/99608f92.fbb1bdd6

27. Louissaint N. Bringing every tool to the COVID-19 fight – what we need now. *Harv Data Sci Rev*. 2021;3(1):2021. doi:10.1162/99608f92.70f8cb98

28. Hamad AF, Walld R, Lix LM, Urquia ML, Roos LL, Wall-Wieler E. Data resource profile: the Manitoba multigenerational cohort. *Int J Epidemiol*. 2022;51(3):e65-e72. doi:10.1093/ije/dyab195

29. Katz A, Enns J, Smith M, Burchill C, Turner K, Towns D. Population data centre profile: The Manitoba Centre for Health Policy. *Int J Popul Data Sci*. 2020;4(2). doi:10.23889/ijpds.v4i2.1131

30. Smith M, Roos LL, Burchill C, et al. Health services data: managing the data warehouse: 25 years of experience at the Manitoba Centre for Health Policy. In: *Data and Measures in Health Services Research*. Springer US; 2015:1-26. doi:10.1007/978-1-4899-7673-4_3-1

31. Roos L. A research registry uses, development, and accuracy. *J Clin Epidemiol*. 1999;52(1):39-47. doi:10.1016/S0895-4356(98)00126-7

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

32. Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. *Soc Sci Med*. 2004;58(11):2231-2241. doi:10.1016/j.socscimed.2003.08.008

33. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011;32(1):91-108. doi:10.1146/annurev-publhealth-031210-100700

34. SAS Institute. SAS Version 15.1. 2023;(9.4).

35. Michelen M, Manoharan L, Elkheir N, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health*. 2021;6(9):e005427. doi:10.1136/BMJGH-2021-005427

36. Hayes LD, Ingram J, Sculthorpe NF. More Than 100 Persistent Symptoms of SARS-CoV-2 (Long COVID): A Scoping Review. *Front Med (Lausanne)*. 2021;8:750378. doi:10.3389/FMED.2021.750378/BIBTEX

37. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nature Reviews Microbiology 2023 21:3*. 2023;21(3):133-146. doi:10.1038/s41579-022-00846-2

38. Silva Filho T, Song H, Perello-Nieto M, Santos-Rodriguez R, Kull M, Flach P. Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Mach Learn*. 2023;112(9):3211-3260. doi:10.1007/s10994-023-06336-7

39. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

40. Lankham I, Slaughter M. Simple and efficient bootstrap validation of predictive models using SAS/STAT® software. In: *Proceedings of the SAS Global Forum 2020 Conferenc*. SAS Institute Inc; 2020. https://support.sas.com/resources/papers/proceedings20/4647-2020.pdf

41. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017;2017:1-14. doi:10.1155/2017/3762651

42. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc*. 2018;113(521):390-400. doi:10.1080/01621459.2016.1260466

43. Zang C, Zhang Y, Xu J, et al. Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative. *Nature Communications 2023 14:1*. 2023;14(1):1-14. doi:10.1038/s41467-023-37653-z

44. Guilcher SJT, Bai YQ, Wodchis WP, Bronskill SE, Kuluski K. An interrupted time series study using administrative health data to examine the impact of the COVID-19 pandemic on alternate care level acute hospitalizations in Ontario, Canada. *Canadian Medical Association Open Access Journal*. 2023;11(4):E621-E629. doi:10.9778/CMAJO.20220086

45. Binka M, Klaver B, Cua G, et al. An elastic net regression model for identifying long COVID patients using health administrative data: a population-based study. *Open Forum Infect Dis*. 2022;9(12):ofac640. doi:10.1093/ofid/ofac640

46. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27(4):601. doi:10.1038/S41591-021-01283-Z

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

47. Perlis RH, Santillana M, Ognyanova K, et al. Prevalence and correlates of long COVID symptoms among US adults. *JAMA Netw Open*. 2022;5(10):E2238804. doi:10.1001/JAMANETWORKOPEN.2022.38804

48. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz J V. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. 2022;22(4):e102-e107. doi:10.1016/S1473-3099(21)00703-9

49. Chevinsky JR, Tao G, Lavery AM, et al. Late Conditions Diagnosed 1-4 Months Following an Initial Coronavirus Disease 2019 (COVID-19) Encounter: A Matched-Cohort Study Using Inpatient and Outpatient Administrative Data-United States, 1 March-30 June 2020. *Clin Infect Dis*. 2021;73(Suppl 1):S5-S16. doi:10.1093/CID/CIAB338

50. Kosowan L, Sanchez-Ramirez DC, Katz A. Understanding symptoms suggestive of long COVID syndrome and healthcare use among community-based populations in Manitoba, Canada: an observational cross-sectional survey. *BMJ Open*. 2024;14(1):e075301. doi:10.1136/BMJOPEN-2023-075301

51. Zhu Y, Mahale A, Peters K, et al. Using natural language processing on free-text clinical notes to identify patients with long-term COVID effects. In: *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. ACM; 2022:1-9. doi:10.1145/3535508.3545555

52. Wong MCS, Huang J, Wong YY, et al. Epidemiology, symptomatology, and risk factors for long COVID symptoms: population-based, multicenter study. *JMIR Public Health Surveill*. 2023;9:1-15. doi:10.2196/42315

53. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. 2019;7(2):e12239. doi:10.2196/12239

54. Luo J, Zhang J, Tang HT, et al. Prevalence and risk factors of long COVID 6–12 months after infection with the Omicron variant among nonhospitalized patients in Hong Kong. *J Med Virol*. 2023;95(6). doi:10.1002/jmv.28862

55. Marra AR, Sampaio VS, Ozahata MC, et al. Risk factors for long coronavirus disease 2019 (long COVID) among healthcare personnel, Brazil, 2020-2022. *Infect Control Hosp Epidemiol*. 2023;44(12):1972-1978. doi:10.1017/ice.2023.95

56. Martínez-Ayala MC, Proaños NJ, Cala-Duran J, et al. Factors associated with long COVID syndrome in a Colombian cohort. *Front Med (Lausanne)*. 2023;10(December):1-9. doi:10.3389/fmed.2023.1325616

57. Morello R, Mariani F, Mastrantoni L, et al. Risk factors for post-COVID-19 condition (Long Covid) in children: a prospective cohort study. *EClinicalMedicine*. 2023;59:101961. doi:10.1016/j.eclinm.2023.101961

58. Rayner DG, Wang E, Su C, et al. Risk factors for long COVID in children and adolescents: a systematic review and meta-analysis. *World Journal of Pediatrics*. 2023;20(2):133-142. doi:10.1007/s12519-023-00765-z

59. United Nations. *Indigenous Peoples and the COVID-19 Pandemic: Considerations*.; 2020. Accessed January 8, 2024. https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2020/04/COVID19_IP_considerations.pdf

/ScholarOne/conversions/7096788968893061517/54507837_File000007_1334835312.docx

60. Greenwood M, de Leeuw S, Lindsay N. Challenges in health equity for Indigenous peoples in Canada. *The Lancet*. 2018;391(10131):1645-1648. Accessed May 15, 2021. https://www-sciencedirect-com.uml.idm.oclc.org/science/article/pii/S0140673618301776

61. Curtice K, Choo E. *Indigenous Populations: Left behind in the COVID-19 Response*. Vol 395.; 2020. doi:10.1016/S0140-6736(20)31242-3

62. Nickel NC, Phillips-Beck W, Enns JE, et al. COVID-19 diagnostic testing and vaccinations among First Nations in Manitoba: A nations-based retrospective cohort study using linked administrative data, 2020–2021. *PLoS Med*. 2024;21(2):e1004348. doi:10.1371/JOURNAL.PMED.1004348

63. Romero-Ibarguengoitia ME, Rodríguez-Torres JF, Garza-Silva A, et al. Association of vaccine status, reinfections, and risk factors with Long COVID syndrome. *Sci Rep*. 2024;14(1):1-9. doi:10.1038/s41598-024-52925-4

64. Komaroff AL, Lipkin WI. ME/CFS and long COVID share similar symptoms and biological abnormalities: road map to the literature. *Front Med (Lausanne)*. 2023;10. doi:10.3389/fmed.2023.1187163

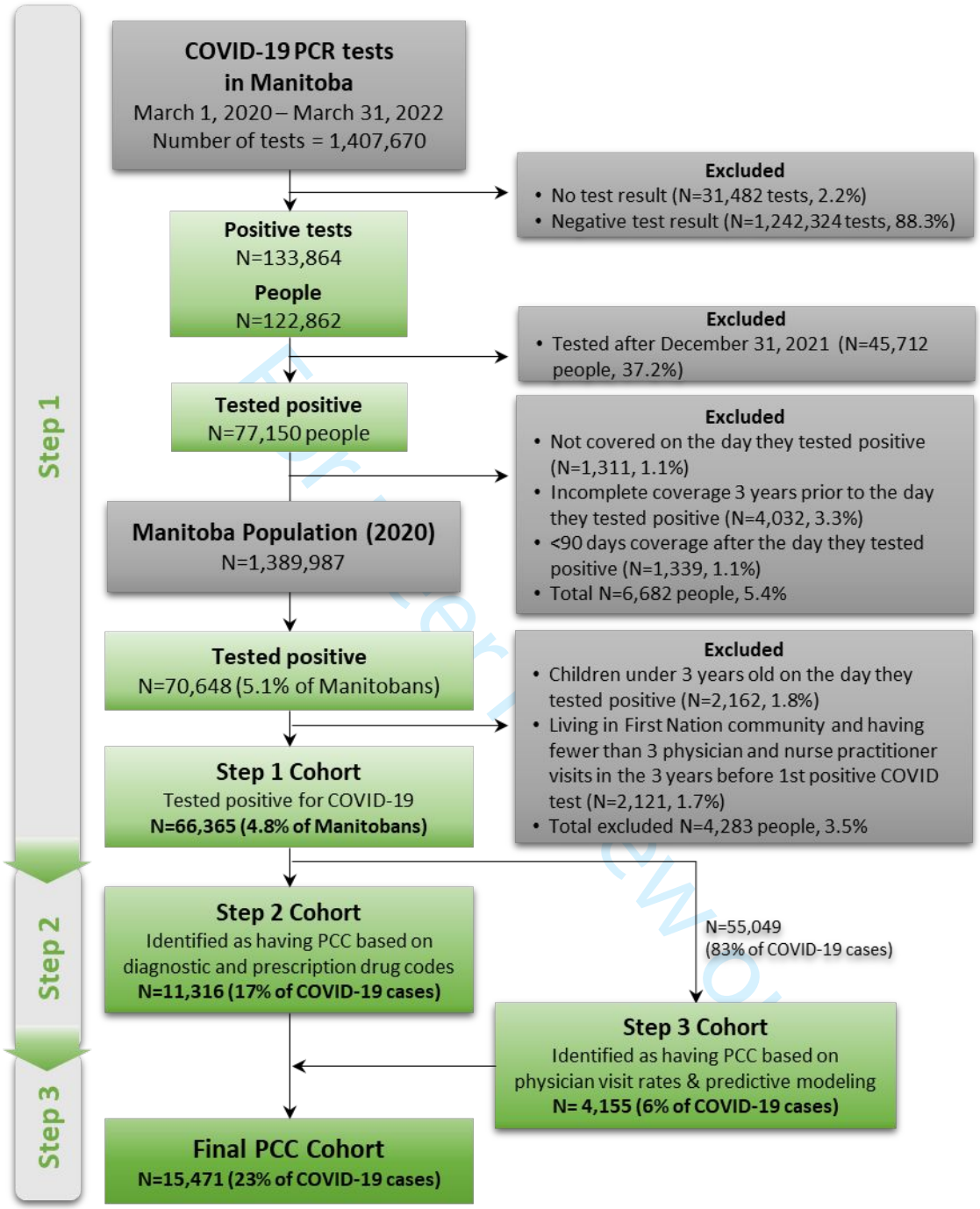
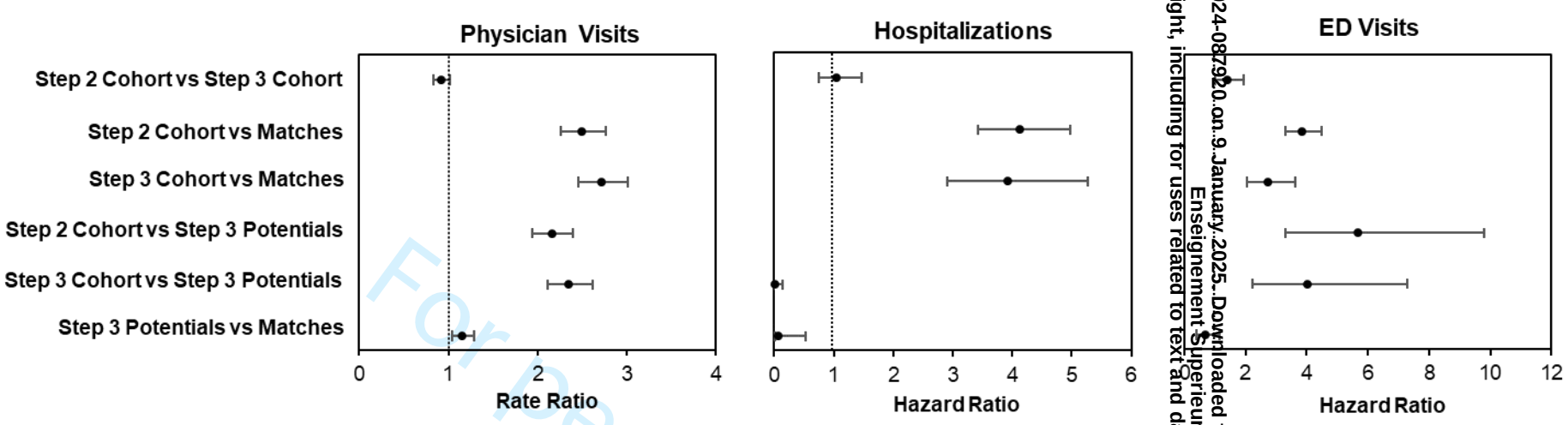


Figure 1. Post-COVID condition cohort development



Step 3 Potentials are the individuals identified as having PCC based on physician visit rates in Step 3.
Matches are the individuals in the HDPS-matched cohort.

Figure 2. Three-way health service use comparison to examine validity of the physician visit predictive modeling approach to identifying post-COVID condition (rate ratios and hazard ratios with 95% confidence intervals)

ED: Emergency department.

Supplemental Material

Table S1. ICD-9-CM codes, ICD-10-CA codes, and ATC codes used to identify post-COVID condition cases in Manitoba (Step 2 of cohort construction)

Disease Category	ICD-9-CM code (out-patient)	ICD-10-CA code (in-patient)	ATC code (prescriptions)
Endocrine, nutritional, and metabolic diseases		E10	
Mental, behavioral, and neurodevelopmental disorders	2962, 2968, 2969, 3004, 3007, 30742, 309, 311	F02, F03, F04, F07, F09, F32, F34, F39, F41, F43, F48, F90	N06A, N06BA
Diseases of the nervous system	3379, 3545, 3548, 3549, 3551, 3555, 3557, 3558, 3559, 3564, 3568, 3589, 3574, 3579	G40, G44, G471, G478, G479, G470, G50, G51, G52, G53, G55, G56, G57, G58, G59, G89, G904, G905, G908, G909, G933	N05B, N05C, N03AE
Diseases of the eye and adnexa		H103, H1042, H1040, H1050, H1089, H109, H162, H53	
Diseases of the ear and mastoid process	3861, 3883, 3887	H813, H814, H818, H819, H90, H920, H931, H932, H93A	
Diseases of the circulatory system	412, 4139, 4148, 4151, 4209, 4254, 4270, 4271, 4272, 4273, 4274, 4275, 4276, 4278, 4279, 428, 4290, 4293, 4298, 4299, 7850, 7851, 431, 432, 433, 434, 4359, 436, 452,4531, 4532, 4533, 4536, 4538, 4539, 4591,	I21, I2602, I30, I313, I314, I32, I40, I44, I45, I46, I47, I48, I49, I50, I5A	C07
Diseases of the respiratory system	4939	J310, J311, J312, J32, J370, J41, J42, J4521, J4522, J4531, J4532, J4541, J4542, J4551, J4552, J4590, J4599, J841	L04AX05, L01EX09, R03, H02A
Diseases of the digestive system		K293, K294, K296, K297, K298, K299, K290, K73, K744, K745, K7469, K7460, K740, K759	
Diseases of the skin and subcutaneous tissue		L25, L26, K299, L308, L309, L63, L65, L66, L984, L988, L989	
Diseases of the musculoskeletal system and connective tissue	7291	M028, M029, M048, M049, M128, M2480, M255, M259, M608, M609, M6281, M6283, M6289, M792, M797	M01

Diseases of the genitourinary system		N18, N394	
Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	7807, 7826	R002, R05, R0600. R0609, R0902, R093, R101, R103, R108, R109, R11, R198, R21, R32, R41, R42, R430, R49, R50, R51, R52, R538, R61, R630, R634, R7, R945,	N02
Injury, poisoning, and certain other consequences of external causes		T1491	
External causes of morbidity		X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83	

ATC: Anatomical Therapeutic Classification; ICD: International Classification of Disease

Table S2. Variables used in our predictive modeling approach to predict post-COVID condition among Manitobans who had tested positive for COVID-19 (Step 3 of cohort construction)

Variables considered in the training models	Variables included in the final model
Sex	**Y**
Age group at first positive COVID-19 test	**Y**
Region of residence (urban/rural)	
Regional Health Authority	**Y**
Income quintile	
SEFI-2 (socioeconomic factor index)	
Identified in medical records as having PCC in the first 90 days after positive COVID-19 test	**Y**
Hospital admission within 14 days of positive COVID-19 test	**Y**
Hospital admission after 14 days of positive COVID-19 test	**Y**
ICU admission within 14 days of positive COVID-19 test	**Y**
Charlson comorbidity index one year prior to positive COVID-19 test	**Y**
Number of COVID-19 vaccine doses received prior to positive COVID-19 test	**Y**
Recent immigrant to Manitoba (within five years of the study period)	
First Nations	

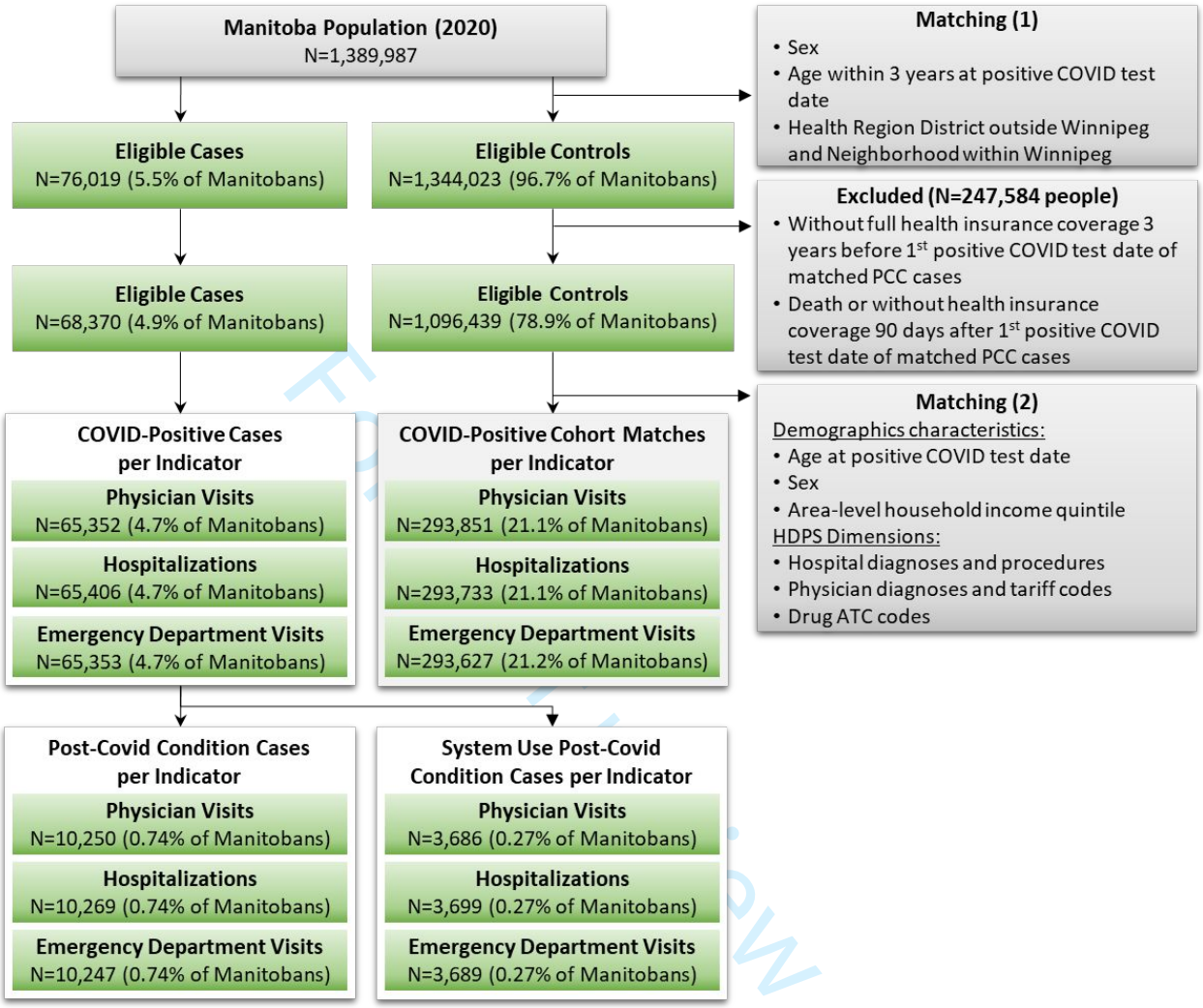


Figure S1. Development of the HDPS-matched cohorts for validation of Step 3 cohort construction

BMJ Open

Identifying people with post-COVID condition using linked, population-based administrative health data from Manitoba, Canada: Prevalence and predictors in a cohort of COVID-positive individuals

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-087920.R1
Article Type:	Original research
Date Submitted by the Author:	21-Nov-2024
Complete List of Authors:	Katz, Alan; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences; Rady Faculty of Health Sciences University of Manitoba, Department of Family Medicine Ekuma, Okechukwu; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Enns, Jennifer; University of Manitoba Faculty of Health Sciences, Manitoba Centre for Health Policy Cavett, Teresa; University of Manitoba Singer, Alexander; Rady Faculty of Health Sciences, University of Manitoba, Family Medicine Sanchez-Ramirez, Diana C.; University of Manitoba; University of Manitoba Keynan, Yoav; University of Manitoba, Medical Microbiology and Infectious Diseases Lix, Lisa; University of Manitoba, Community Health Sciences Walld, Randy; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Yogendran, Marina; Manitoba Centre for Health Policy Nickel, Nathan C.; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Urquia, Marcelo; Rady Faculty of Health Sciences University of Manitoba, Department of Community Health Sciences Star, Leona; First Nations Health and Social Secretariat of Manitoba Olafson, Kendiss; Rady Faculty of Health Sciences University of Manitoba, Department of Internal Medicine Logsetty, Sarvesh; University of Manitoba, Departments of Surgery and Psychiatry Spiwak, Rae; University of Manitoba, Department of Surgery Waruk, Jillian; First Nations Health and Social Secretariat of Manitoba Matharaarachichi, Surani; University of Manitoba, Department of Statistics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health services research, Infectious diseases, Public health

Keywords:	SARS-CoV-2 Infection, COVID-19, Post-Acute COVID-19 Syndrome, Epidemiology < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Title:

Identifying people with post-COVID condition using linked, population-based administrative health data from Manitoba, Canada: Prevalence and predictors in a cohort of COVID-positive individuals

Authors and Affiliations:

Alan Katz[1,2], Okechukwu Ekuma[1], Jennifer E. Enns[1], Teresa Cavett[2], Alexander Singer[2], Diana C. Sanchez-Ramirez[3], Yoav Keynan[4], Lisa M. Lix[1], Randy Walld[1], Marina S. Yogendran[1], Nathan Nickel[1], Marcelo L. Urquia[1], Leona Star[5], Kendiss Olafson[4], Sarvesh Logsetty[6], Rae Spiwak[6], Jillian Waruk[5], Surani Matharaarachichi[7]

1. Manitoba Centre for Health Policy, Department of Community Health Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
2. Department of Family Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
3. College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
4. Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
5. First Nations Health and Social Secretariat of Manitoba, Winnipeg, Canada
6. Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
7. Department of Statistics, University of Manitoba, Winnipeg, Canada

Corresponding Author:

Dr. Alan Katz
Manitoba Centre for Health Policy
University of Manitoba
408-727 McDermot Ave
Winnipeg, Manitoba
R3E 3P5
Alan.Katz@umanitoba.ca

Abstract

Objective: Many individuals exposed to SARS-CoV-2 experience long-term symptoms as part of a syndrome called post-COVID condition (PCC). Research on PCC is still emerging but is urgently needed to support diagnosis, clinical treatment guidelines, and health system resource allocation. In this study, we developed a method to identify PCC cases using administrative health data, and report PCC prevalence and predictive factors in Manitoba, Canada.

Design: Cohort study.

Setting: Manitoba, Canada.

Participants: All Manitobans who tested positive for SARS-CoV-2 during population-wide PCR testing from March 2020 to December 2021 (n=66,365), and were subsequently deemed to have PCC based on ICD-9/10 diagnostic codes and prescription drug codes (n=11,316). Additional PCC cases were identified using predictive modelling to assess patterns of health service use, including physician visits, emergency department visits, and hospitalization for any reason (n=4,155).

Outcomes: We measured PCC prevalence (% PCC cases among Manitobans with positive tests) and identified predictive factors associated with PCC (odds ratios with 95% confidence intervals, adjusted for socio-demographic and clinical characteristics).

Results: Among 66,365 Manitobans with positive tests, we identified 15,471 (23%) as having PCC. Being female (aOR 1.64, 95% CI 1.58-1.71), age 59 or older (age 60-79 aOR 1.33, 95% CI 1.25-1.41; age 80+ aOR 1.62, 95% CI 1.46-1.80), hospitalized within 14 days of COVID-19 infection (aOR 1.95, 95% CI 1.80-2.10), and having a Charlson Comorbidity Index of 1+ (aOR 1.95, 95% CI 1.78-2.14) were predictive of PCC. Receiving 1+ doses of the COVID-19 vaccine (1 dose aOR 0.80, 95% CI 0.74-0.86; 2 doses aOR 0.29, 95% CI 0.22-0.31) decreased the odds of PCC.

Conclusions: This data-driven approach expands our understanding of the prevalence and epidemiology of PCC and may be applied in other jurisdictions with population-based data. The study provides additional insights into risk and protective factors for PCC to inform health system planning and service delivery.

Keywords: SARS-CoV-2; post-COVID condition; long COVID; administrative health data; data linkage; epidemiology; predictive modeling; Canada

Funding

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

The study received funding from a Canadian Institutes of Health Research Operating Grant: Emerging COVID-19 Research Gaps and Priorities (Grant No. GA7-1777787) awarded to Alan Katz in 2020. The results and conclusions are those of the authors and no official endorsement by the funder is intended or should be inferred.

Competing Interests

The authors declare that they have no competing interests.

Strengths and Limitations

- The Manitoba Population Research Data Repository holds an extensive collection of administrative datasets that are population-based and linkable across domains at the individual and family levels, making it a valuable resource for population health research.
- We had access to multiple sources of relevant data (COVID-19 test and vaccination data, physician billing claim codes, prescribed medications, and physician visit and hospitalization patterns) from which to triangulate our approach and identify people with post-COVID condition.
- Despite the strengths of this approach, we cannot be certain that all of the symptoms and health service use patterns we identified as post-COVID condition were all directly related, since providers and patients may not have recognized their symptoms as being post-COVID condition at the time of the visit, or may have been seeking medical help for symptoms that were not actually post-COVID condition.
- Using natural language processing to incorporate free text from physicians' clinical notes into our analysis would provide additional context and nuance to our administrative data-driven approach, and may be an option for future research in this area.

Introduction

The SARS-CoV-2 pandemic created significant societal disruption across the globe, with over 774 million confirmed cases and more than 7 million deaths as of early 2024.[1] While COVID-19 is primarily considered an acute illness, some individuals experience long-term symptoms in a syndrome referred to as 'long COVID' or 'post-COVID condition' (PCC). More than 200 symptoms have been reported to occur in the weeks following SARS-CoV-2 infection, including shortness of breath, cognitive dysfunction, and fatigue, as well as multiple conditions that impact daily functions, such as neurological and psychiatric disorders, hypertension, respiratory difficulties, heart palpitations, chest pain, myalgia, and digestive disorders.[2–4]

Our current understanding of PCC is mainly derived from a series of clinical cohort, cross-sectional, and survey-based studies.[5–11] Estimates of PCC prevalence vary, with 32-87% of COVID-19 survivors reporting persistent or new symptoms two to three months after acute infection.[12] Many individuals with PCC are young, previously healthy, and had mild forms of COVID-19 that did not require hospitalization.[13] There appears to be higher prevalence of PCC in particular groups, including females and individuals aged 35-45.[13–15] SARS-CoV-2 infection has also been shown to disproportionately impact many communities who are lower income, have higher rates of comorbidities, live in crowded or inadequate housing conditions, and have poorer access to health services than the general population.[16] This could mean a higher prevalence of PCC in, for example, Indigenous populations or immigrant groups.[17,18]

PCC can have a tremendous impact on quality of life and has been shown to substantially increase health service use.[19–21] The underlying mechanisms of PCC are now presumed to be mediated by the immune system.[22,23] The link between complement dysregulation and PCC explains the three

principal mechanisms: immune dysregulation, immune priming, and microvascular blood clotting[23]; as well, the activation of cytomegalovirus- and Epstein-Barr virus-specific antibodies may also contribute to the pathophysiology of PCC.[23,24]

Despite PCC becoming better understood, several important clinical questions remain. Due to the emerging nature of the illness, the large number of associated symptoms, and the diverse methodologies used thus far to study PCC, there is currently no consensus on the true clinical definition of PCC. While the World Health Organization (WHO), the Government of Canada, the National Institute for Clinical Excellence, and the Centres for Disease Control and Prevention have each proposed different clinical definitions, these may continue to undergo refinement as the understanding of PCC and its pathophysiology evolves. In the meantime, an analytic approach for identifying PCC in the general population would be a significant asset, since it would support the investigation of risk characteristics, medium- to long-term health outcomes, and recovery trajectories of those with PCC. This evidence is critical for informing the development of clinical treatment guidelines and health system resource planning and would also aid in the development of COVID-related health policy worldwide.

Several editorials have called for data science-based solutions to address some of the challenges the pandemic has presented.[25–27] Administrative data are a powerful resource for researchers, clinicians, patients, and health system decision-makers, and are currently being used to help us understand COVID-19 and its impact across various populations. An administrative data approach has the advantage of capturing a whole population of interest, thereby limiting selection bias and loss to follow-up, and because administrative data include community-based care, they allow for the inclusion of sub-populations with less severe disease that may not be present in hospital or critical care studies. Research using administrative data to examine prevalence of poorly defined

syndromes in the US population, for example, post-treatment Lyme disease syndrome[28] and (most recently) PCC[29], has demonstrated the utility of such an approach. In this study, we constructed a PCC cohort using administrative health data from a population-based data repository in Manitoba, Canada, and described PCC risk and protective factors in the COVID-positive population.

Methods

Study Setting

Manitoba is a province with approximately 1.4 million residents in the geographical centre of Canada. The provincial single-payer healthcare system covers over 99% of Manitoba residents, excluding only individuals who are federally insured (e.g., those incarcerated in federal prisons, members of the military, and some First Nations and Inuit populations).[30,31] The provincial health insurance coverage includes all hospitalizations, medically necessary physician services, and prescription drug dispensations by Manitoba pharmacies.

During the study period (March 2020 to December 2021), the total number of people residing in Manitoba was 1,465,704. In March 2020, the population was 1,385,424 and in December 2021 it was 1,405,498, with an overall increase of 20,074 people, a growth rate of 1.4% during the study period. There were 30,237 births (2.06%), 21,721 deaths (1.48%), 52,243 people who moved into Manitoba or initiated health insurance coverage for a reason other than birth (3.56%), and 40,764 people who moved out of the province or ceased health insurance coverage for a reason other than death (2.78%).

Data Sources

The Manitoba Population Research Data Repository at the Manitoba Centre for Health Policy comprises over 100 databases of whole-population, individual-level administrative data from the health, social services, legal, and education systems in Manitoba. All Repository databases are de-identified (names and addresses removed), but they are linkable at the individual and family level using a scrambled Patient Health Identification Number attached to each record and to a central population registry.[31,32] The Repository data have been used extensively in research and the validity of the databases to examine population health has been well documented.[33–35] The databases used in this study are presented in **Table 1**. Both the hospital discharge abstract data and the medical claims data have been shown to be comprehensive for the population of Manitoba.

Table 1. Databases and Key Variables from the Manitoba Population Research Data Repository

Database	Key Variables or Information
Manitoba Health Insurance Registry	Biological sex, date of birth, and postal code for the population of Manitoba
COVID-19 Laboratory Testing and Results	COVID-19 PCR test information (date, test result)
COVID-19 Vaccinations	COVID-19 vaccinations (date, type of vaccine, number of doses)
Hospital Discharge Abstracts	ICD-10 diagnosis and procedure codes from in-patient hospital stays
Hospital Admission, Discharge, and Transfer Database	Data on hospital and ICU admission, discharge, and transfer
Medical Claims/Medical Services	Billing claims from primary care clinics (physicians and nurse practitioners) and specialist physicians, including ICD-9 diagnosis and procedure codes
Drug Program Information Network	All prescription medications dispensed by community retail pharmacies
Emergency Department Information System	Data on emergency department visits
Immigration, Refugees, and Citizenship Canada (IRCC) Permanent Resident Database	Data on people who immigrated to Manitoba
Manitoba First Nations Research File	People registered as First Nations according to the Indian Act of Canada

ICD: International Classification of Disease; ICU: intensive care unit; PCR: Polymerase Chain Reaction

Patient/Public Involvement

At least one member of the research team was an individual who was experiencing PCC at the time of the study. The study conception, design, conduct, and interpretation were informed by their personal experiences with PCC and as a family physician. They are also listed as an author on the study.

Ethics and Privacy

Ethics approvals were granted by the University of Manitoba's Health Research Ethics Board (#HS25090 H2021:279) and the Health Information Research Governance Committee at the Manitoba First Nations Health and Social Secretariat (2020). The study protocol was reviewed by the Manitoba Government Health Information Privacy Committee (HIPC #2021/2022–18).

Constructing the PCC Cohort

The cohort construction comprised three key steps, an overview of which is presented in **Figure 1**. All cohort construction steps and analyses were completed in SAS/STAT® software version 15.1.[36]

Step 1. Identify the population at risk for PCC.

We began by identifying all Manitobans at risk for developing PCC, i.e. those with a positive COVID-19 test result from March 1, 2020 to December 31, 2021 (Step 1 cohort). People in the cohort were followed from their inclusion in the COVID-positive cohort for at least 90 days and for at most 365 days after index date, or until they were lost to follow-up (due to death or a move out of province) or the end of study date (June 30, 2021). The province of Manitoba provided whole-population COVID-19 PCR testing from March 2020 to December 2021, thus we were able to include information on the *alpha*, *delta,* and early *omicron* waves of infection. The province's centralized testing approach was discontinued on December 31, 2021.

Step 2. Use diagnosis codes and prescription drug codes to narrow to those with PCC.

From the rapidly evolving literature on COVID-19,[3,20,37–39] we gathered information on symptoms commonly associated with PCC and medications used to treat these symptoms. We determined the corresponding International Classification of Disease (ICD)-9 (out-patient) and ICD-10 (in-hospital) diagnosis and procedure codes and Anatomical Therapeutic Chemical (ATC) codes (**Table S1**), and used these to narrow the cohort to only people who reported new symptoms or received a new prescription related to PCC symptoms 90 days or more after their positive COVID-19 test date (Step 2 cohort). We excluded anyone who reported the same symptoms in the three years before their positive COVID-19 test. However, given the broad range of symptoms associated with COVID-19, we knew this step likely wouldn't be exhaustive, because individuals with PCC might not have reported new symptoms (many are common to other illnesses) and might not have received a new prescription if their symptoms were ones they had experienced previously. Thus, we used another strategy to identify additional individuals in the Step 1 cohort likely to have experienced PCC but not captured in Step 2.

Step 3. Identify additional individuals with PCC based on predictive modeling.

In this final step, we used a predictive modeling approach to identify individuals in Manitoba likely to have PCC based on their physician visit rates. This approach allowed us to incorporate socio-demographic and health service use variables into the comparison between the Step 2 cohort and the rest of the population of Manitoba using a High Dimensional Propensity Score (HDPS).

First, in the PCC cohort from Step 2, we calculated the physician visit rate from the 91st day after their first positive COVID-19 test date to March 31, 2022 (PCC period), and compared it to their visit rate from April 1, 2019 to March 31, 2020 (pre-COVID period). We also compared physician visit rates for these two periods among those that tested positive for COVID-19 in Step 1 but were not part

of the Step 2 cohort. In this latter group, those with visit rates the same or greater than those of the Step 2 cohort were added to the final cohort. Then, we trained several different regression model builds (step-wise, best subset, and lasso) to determine the best fit. Model selection was done by considering the Brier score, the Schwarz Bayesian Criterion, and the C-statistic,[40,41] and the selected model was then internally validated by bootstrapping 500 samples, yielding an optimism-corrected C-statistic of 0.69387.[42] This method ensured that we used all the data we had available in the prediction process. The selected model was then re-fitted to the entire group of individuals who tested positive for COVID-19 (Step 1 cohort) to determine the predicted probability of any of those individuals having PCC. The list of variables considered in these models and the ones that were used in the selected model are shown in **Table S2.**

One of the notable challenges in predictive modeling is determining the optimal cut-off with which to classify individuals as predicted to have the outcome. We tested several methods to obtain the highest optimal cut-point for a more conservative classification,[43] and then classified the Step 1 cohort into those who were and were not predicted to have PCC.

Validating Step 3 of Cohort Construction

Our efforts to identify individuals with PCC in Step 3 were validated in a three-way comparison: we compared three health service use indicators (physician visit rates, hospitalizations, and emergency department visits) in the Step 2 cohort, the additional individuals identified in Step 3, and a group matched on an HDPS that encompassed both socio-demographic characteristics and health service use data dimensions. We hypothesized that the Step 3 cohort would have health service use patterns similar to the Step 2 cohort.

High Dimensional Propensity Score Matching

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Following the method developed by Li et al.,[44] we hard-matched on age, sex, and region of residence (urban/rural) at the time of first positive COVID-19 test using a ratio of 25:1 and calculated standardized differences on these variables. Then, we generated the HDPS using demographic variables (age, sex, region of residence, income quintile) and health service use data dimensions (hospital diagnosis, hospital procedures, physician diagnosis, physician tariff codes, ATC codes for prescription drugs) (**Figure S1**). The HDPS models were stratified by sex and run separately for three health service use indicators: physician visits, hospitalizations, and emergency department visits. To create the final matched group, the five closest matches to the HDPS in each pool of 25 were selected. The case count was N=66,365. A caliper of 0.2× the pooled standard deviation of the logit (p-score) was applied to avoid selecting pairs with unsuitably large differences in p-score.

In terms of completeness, for physician visits, 52,960 cases had 5 matches (79.80%), 12,392 cases had 1-4 matches (18.67%) and 1,013 cases had no match (1.53%). For hospitalizations, 52, 867 cases had 5 matches (79.66%), 12,539 cases had 1-4 matches (18.89%) and 959 cases had no match (1.45%). For emergency department visits, 52, 885 cases had 5 matches (79.69%), 12,468 cases had 1-4 matches (18.79%) and 1,012 cases had no match (1.52%).***Statistical Analysis***

The results of the three-way comparison used to validate the Step 3 cohort creation process are shown in **Figure 2**. We examined physician visit rates using negative binomial distribution with the log of person-days as the offset, obtaining rate ratios and 95% confidence intervals, and we conducted time-to-event analyses of first hospitalizations and first emergency department visits after index date, obtaining hazard ratios and 95% confidence intervals. The time-to-event approach was used for the latter two indicators because they were rare events.

Predictors of Post-COVID Condition: Univariate and Multivariate Analyses

We used univariate and multivariate logistic regression modeling to identify socio-demographic and clinical variables that were predictive or protective of PCC, producing odds ratios and 95% confidence intervals. Each of the univariate models had PCC as the dependent variable and one of the following as the independent variable: sex (male/female); age group (0-18, 19-29, 30-59, 60-79, 80+); region of residence (urban/rural); income quintile (Q1-Q5); Regional Health Authority (Winnipeg, Interlake-Eastern Regional Health Authority, Prairie Mountain Health, Southern Health/Santé-Sud, Northern Health Region); hospitalization within 14 days of positive COVID-19 test for any reason (Y/N); number of COVID-19 vaccine doses before first positive COVID-19 test (0-3+); comorbidity prior to the start of the study period (Charlson Comorbidity Index, 0-3+); immigrant to Manitoba within the last five years (Y/N); and First Nations (Y/N). The multivariate models included all of the variables in the list above.

Results

PCC Prevalence

As shown in **Figure 1**, we identified 66,365 individuals who tested positive for COVID-19 from March 1, 2020 to December 31, 2021 in Step 1. Among these individuals, 11,316 (17.1%) were identified as having PCC based on their diagnostic codes and prescription drug codes in Step 2. Physician visit patterns for the remaining 55,049 individuals who tested positive for COVID-19 were then further examined in Step 3. The pre-COVID vs PCC period physician visit rate ratio for the Step 2 cohort (N=11,316) was 1.75; among the other 55,049 individuals who tested positive for COVID-19, we identified 4,155 (7.5% of 55,049) individuals as likely having PCC in Step 3 based on their physician visit rate and/or the predictive modeling. This group was added to the Step 2 cohort for a total of

15,471 PCC cases. Thus, we determined the prevalence of PCC to be 23.3% in the COVID-positive population of Manitoba.

Rate ratios for physician visits and hazard ratios for hospitalizations and emergency department visits were calculated as part of the validation process for constructing the final PCC cohort to determine whether there were significant differences in health service use at each step. Shown in **Figure 2**, these data confirm that the rates of health service use were either the same or slightly higher between the Step 2 and Step 3 cohorts.

Predictors of Post-COVID Condition

In the final PCC cohort, we conducted univariate and multivariate analyses to determine the odds of developing PCC based on the cohort's socio-demographic characteristics. The univariate analyses are presented in **Table 2** as odds ratios and 95% confidence intervals**.** Female individuals and those over the age of 59 had higher odds of developing PCC, as did people in lower income brackets, people with comorbidities (measured by the Charlson Comorbidity Index), and people who were hospitalized within 14 days of a positive COVID-19 test. First Nations also had higher odds of developing PCC than other Manitobans, whereas immigrants who arrived in Manitoba within the last five years had lower odds. The COVID-19 vaccine was a protective factor and a higher number of vaccine doses (from 0-3+) conferred greater protection against PCC.

The multivariate analyses are shown in **Table 3**. Even after adjusting for age, sex, income, and region of residence, many of the same factors were predictive of developing PCC, but a few discrepancies between the univariate and the multivariate findings are notable: in the multivariate findings, lower income was no longer a significant predictor of PCC, nor was living in the more remote regions of the province (e.g., the Northern Health Region).

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Table 2. Predictive characteristics for developing PCC in the population of Manitoba, Canada

Univariate odds ratios and 95% confidence intervals

	Odds Ratio	95% CI
Sex		
Male (ref)		
Female	1.60	1.54, 1.66
Age group (years)		
0-18	0.40	0.38, 0.42
19-29	0.87	0.83, 0.91
30-59 (ref)		
60-79	1.54	1.46,1.63
80+	2.20	2.01, 2.41
Region of residence		
Urban (ref)		
Rural	0.92	0.89, 0.96
Income quintile		
Q1 (lowest)	1.32	1.24, 1.40
Q2	1.24	1.17, 1.32
Q3	1.14	1.07, 1.22
Q4	1.10	1.03, 1.17
Q5 (highest) (ref)		
Not found	2.13	1.78, 2.56
Missing	1.72	1.43, 2.07
Regional Health Authority		
Winnipeg Regional Health Authority (ref)		
Interlake-Eastern Regional Health Authority	0.86	0.80, 0.92
Northern Health Region	1.11	1.05, 1.18
Southern Health-Santé Sud	0.80	0.76, 0.84
Prairie Mountain Health	0.98	0.91, 1.04
Charlson Comorbidity Index		
0 (no comorbidity) (ref)		
1	1.79	1.71, 1.88
2	2.14	1.98, 2.33
3+ (higher comorbidity)	3.20	2.94, 3.47
Hospitalization within 14 days of positive COVID-19 test	3.28	3.05, 3.52
Number of COVID-19 vaccine doses before positive COVID-19 test		
0 doses (ref)		
1 dose	0.93	0.87, 0.99
2 doses	0.34	0.32, 0.36
3+ doses	0.35	0.30, 0.40
Recently immigrated to Manitoba (within 5 years of the study period)	0.84	0.77, 0.91
First Nations	1.20	1.15, 1.25

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Table 3. Predictive characteristics for developing PCC in the population of Manitoba, Canada
Multivariate odds ratios and 95% confidence intervals

	Odds Ratio	95% CI
Sex		
Male (ref)		
Female	1.64	1.58, 1.71
Age group (years)		
0-18	0.37	0.35, 0.39
19-29	0.92	0.87, 0.97
30-59 (ref)		
60-79	1.33	1.25, 1.41
80+	1.62	1.46, 1.80
Region of residence		
Urban (ref)		
Rural	0.93	0.81, 1.06
Income quintile		
Q1 (lowest)	1.00	0.94, 1.07
Q2	0.99	0.93, 1.06
Q3	1.01	0.95, 1.08
Q4	0.99	0.92, 1.06
Q5 (highest) (ref)		
Not found	0.62	0.51, 0.76
Unknown	1.15	0.94, 1.40
Regional Health Authority		
Winnipeg Regional Health Authority (ref)		
Interlake-Eastern Regional Health Authority	0.90	0.78, 1.05
Northern Health Region	1.12	0.97, 1.30
Southern Health-Santé Sud	0.80	0.69, 0.92
Prairie Mountain Health	1.11	1.00, 1.23
Charlson Comorbidity Index		
0 (no comorbidity) (ref)		
1	1.40	1.32, 1.47
2	1.46	1.33, 1.60
3+ (higher comorbidity)	1.95	1.78, 2.14
Hospitalization within 14 days of positive COVID-19 test	1.95	1.80, 2.10
Number of COVID-19 vaccine doses before positive COVID-19 test		
0 (ref)		
1	0.80	0.74, 0.86
2	0.29	0.22, 0.31
3+	0.26	0.22, 0.30
Recently immigrated to Manitoba (within 5 years of the study period)	0.88	0.80, 0.96
First Nations	1.12	1.06, 1.18

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Discussion

Our study aimed to develop an administrative data-driven analytic approach to identifying PCC cases in the general population. We identified PCC cases by first identifying individuals with a positive COVID-19 PCR test, then used diagnostic codes, prescription drug codes, and patterns of health service use to develop a cohort of 15,471 individuals with PCC, which represents a prevalence of 23.3% in the segment of the Manitoba population who tested positive for COVID-19. In this cohort, individuals who were female, age 59+, First Nations, had one or more comorbidity, or were hospitalized within two weeks of a positive COVID-19 test had higher odds of developing PCC than other Manitobans. Vaccination for COVID-19 (one or more doses) was protective against PCC.

In the literature, we observed that others using administrative data to examine PCC prevalence generally sought to define PCC using a combination of health provider claims, hospital discharge abstracts, and/or electronic health records.[29,45,46] A study from British Columbia, Canada, used data from patients attending a PCC-specific clinic and patients with an ICD-10-coded hospital admission for PCC, and then identified other potential PCC cases with similar demographics, pre-existing conditions, and COVID-19 symptoms using an elastic net penalized logistic regression model; they reported a PCC prevalence of 18% in the COVID-positive population.[47] The authors acknowledged that they likely underestimated the prevalence of PCC since their analyses were based on the characteristics of severely impacted patients who were either hospitalized or referred to the specialized PCC clinic. However, they had the advantage of established ICD-10 codes, whereas we did not have a specific code for PCC at the time of our study. Other possible reasons for the difference in estimates between ours and others' are the lag time between a positive PCR test and the development of PCC symptoms (some authors used 30-60 days,[48,49] while we used the WHO

definition of 90 days[50]) and the severity of illness in the starting population (e.g., PCC among people hospitalized for acute COVID-19[51]).

The major challenge we encountered in developing our analytic approach was that people experiencing PCC symptoms were not readily identified in the Manitoba Population Research Data Repository. Without access to a specific ICD code for PCC, we instead used surrogate markers (physician billing claim codes, prescribed medications, and physician visit and hospitalization patterns), but we cannot be certain that these were all directly related to PCC. In addition, outpatient billing practices in Manitoba capture only one main diagnosis, potentially limiting the information we had available to identify PCC cases. As well, a possible lack of awareness or understanding of PCC at the time of providing the clinical care and completing the billing claims could have resulted in providers not screening for persistent symptoms and patients not seeking medical help specifically for PCC, especially if symptoms were mild. In an online survey of Manitobans in 2022, 62.6% indicated that PCC symptoms impacted their daily lives, but only 50.8% sought healthcare (among these, 65.9% primary care, 15.2% emergency department, and 32.0% specialist or therapist).[52] To account for these limitations, we relied on the breadth of data available in the repository, drawing from multiple linked databases and using analytic techniques to assemble the PCC cohort. At each analytic step, the research team discussed how the results aligned with their clinical observations and with the literature, and this discussion informed the next step.

A second challenge in our approach was that we weren't certain which variables should be included in the predictive models to identify potential PCC cases, and it's possible that our approach overestimated the prevalence of PCC by including false positives in the cohort. As more research on PCC becomes available, our model could be adjusted to include additional variables, thus increasing sensitivity, and leading to a more refined classification of PCC cases. It would also likely be helpful to

include clinical notes in our administrative data definition of PCC. Our structured administrative claims data currently lack the context that would be provided by linked clinical notes from physician visits. Recent advances in natural language processing would facilitate the incorporation of free text into analyses of structured administrative data.[53–55]

With respect to PCC predictors, we identified several socio-demographic factors that were associated with higher odds of developing PCC. Several other studies have reported similar findings, particularly the higher odds of PCC among female and/or older individuals, people who were severely ill after contracting acute COVID-19 (e.g., required hospitalization), and people with certain types of comorbidities.[54,56–60] Higher odds of PCC among First Nations has not been reported previously, although there is ample published research showing how colonial practices and policies have disadvantaged First Nations and other Indigenous populations over many generations and placed them at risk for poor health outcomes, particularly in a public health crisis like the SARS-CoV-2 pandemic.[61–63] A recent study reports on the public health measures implemented in Manitoba, how First Nations acted to protect their communities from COVID-19, and their success in advocating for vaccine prioritization.[64] Immigrants having lower odds of developing PCC than other Manitobans can perhaps be explained by the healthy immigrant effect,[18] although examining sub-groups of immigrants would likely provide more insight into this finding. In our current study, vaccination was found to be protective against PCC, and this finding is generally supported by existing literature,[54,57,59,65] although a 2023 study by Luo et al[56] did not find a significant decrease in PCC risk with more vaccine doses, possibly because the study examined a later time period than ours, where most individuals already had 3+ doses of the COVID-19 vaccine. Finally, while some of our analyses (data not presented here) indicate an income gradient in positive COVID-19 tests, this gradient is not evident in the PCC population identified in our study.

The method we present here for estimating the prevalence of PCC using linked administrative data is widely generalizable to other jurisdictions with access to health system data and can be used to follow long term outcomes in the PCC population. It will serve as the basis for further research (ours and others') to follow the natural history and health outcomes of PCC. In particular, we plan to examine PCC outcomes in Indigenous and immigrant populations in more detail. Our findings are being shared widely within our research team's networks of healthcare providers and planners to support the resourcing and implementation of health services for patients living with PCC. As well, the ongoing discourse on the pathogenetic mechanisms of PCC draws many parallels with other post-viral conditions,[66] and as such, our approach could inform future research to better understand the prevalence and course of these conditions. A few limitations of our study approach warrant mention. By definition, our study cohorts only included those with a positive COVID-19 test result. Access to COVID-19 tests may not have been equitable across all of Manitoba for the entire study period, especially during early stages of the pandemic. It is possible that individuals with minor symptoms chose not to complete a test and thus our method selects for more severe disease. We also assumed that patients hospitalized with COVID-19 had more severe disease than those who were not hospitalized. Finally, the constant genetic variation and the evolution of new variants and sub-variants and the changing application of testing and vaccination modalities limits our ability to describe the current prevalence of COVID-19 and PCC. However, as more data become available, we could conduct time series analyses based on the emergence of new variants to address this limitation.

Conclusion

Population-wide administrative data have the potential to address important gaps in our knowledge of emerging health conditions, and can provide nuanced information for clinical and

health system recovery and prevention efforts in the aftermath of a global public health crisis like the SARS-CoV-2 pandemic. Their value may be further enhanced by the use of natural language processing to link unstructured clinical data to existing repositories, which would provide additional context to structured administrative data.

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

List of Abbreviations

ATC	Anatomical Therapeutic Chemical
HDPS	High Dimensional Propensity Score
ICD-9	International Classification of Disease (out-patient)
ICD-10	International Classification of Disease (in hospital)
PCC	Post-COVID Condition
PCR	Polymerase Chain Reaction
WHO	World Health Organization

Acknowledgements

The authors acknowledge the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under project #2021-033 (HIPC/PHRPC #2021/2022-18). 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 databases provided by Manitoba Health; Statistics Canada; Shared Health Manitoba; Immigration, Refugees, and Citizenship Canada; First Nations Health and Social Secretariat of Manitoba; and Manitoba Primary Care Research Network. The results and conclusions are those of the authors and no official endorsement by Manitoba Health or other data providers is intended or should be inferred.

We would also like to thank Annshirley Aba Afful, Quinn Anhorn, Dana Kliewer, Katie Kitchen, Marlee Mayer, Saman Muthukumaranra, and Gillian Martens for their contributions to analysis planning and manuscript writing.

Author Contributions

AK, JEE, TC, AS, DCS-R, YK, LML, NN, MLU, and LS were responsible for research conception; all of these authors as well as KO, SL, and RS contributed to study design. OE, RW, and MSY conducted the data analyses. All authors, including JW and SM, contributed to the interpretation of the analytic findings. JEE and AK drafted the manuscript. All authors reviewed the draft manuscript for intellectual content and gave approval for publication. All authors agree to be accountable for all aspects of the work. Alan Katz is responsible for the overall content as guarantor.

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Figure Legends

Figure 1. Post-COVID condition cohort development

Figure 2. Three-way health service use comparison to examine validity of the physician visit/predictive modeling approach to identifying post-COVID condition (rate ratios and hazard ratios with 95% confidence intervals)**.** ED: Emergency department.

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

Data Sharing Statement

The source data used in this study were originally collected during the routine administration of health and social services in Manitoba, and were provided to the Manitoba Centre for Health Policy (MCHP) for secondary use in research under specific data sharing agreements between the data trustees and MCHP. The data are approved for use at MCHP only. They are not owned by the researchers or by MCHP and cannot be deposited in a public repository. To review source data specific to this article or project, interested parties should contact the MCHP Repository Access & Use team at MCHP.Access@umanitoba.ca. The team will then facilitate data access by seeking the consent of the original data holders and the required privacy and ethics review bodies. Research studies using First Nations data require ethics approval from the Health Information Research Governance Committee at the First Nations Health and Social Secretariat of Manitoba (https://www.fnhssm.com/hirgc), and we comply with their policies for data access, linkage and sharing. For inquiries about accessing Manitoba First Nations data, please contact info@fnhssm.com.

/ScholarOne/conversions/5287328289734763276/57346981_File000011_1409399313.docx

References

1. World Health Organization. COVID-19 Epidemiological Update 2024. Coronavirus disease (COVID-19) weekly epidemiological updates and monthly operational updates. 2024. Accessed February 8, 2024. https://www.who.int/publications/m/item/covid-19-epidemiological-update---19-january-2024

2. Centers for Disease Control and Prevention. Post-COVID/Long COVID. 2020. Accessed March 29, 2021. https://www.idsociety.org/covid-19-real-time-learning-network/disease-manifestations--complications/post-covid-syndrome/

3. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine*. 2021;38:101019. doi:10.1016/J.ECLINM.2021.101019

4. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry*. 2021;8(5):416-427. doi:10.1016/S2215-0366(21)00084-5

5. Chen X, Norris C, Whitten T, Ho C, Mann B, Bakal J. Symptomology following COVID-19 among adults in Alberta, Canada: an observational survey study. *BMJ Open*. 2024;14(3):e078119. doi:10.1136/bmjopen-2023-078119

6. Cohen J, van der Meulen Rodgers Y. An intersectional analysis of long COVID prevalence. *Int J Equity Health*. 2023;22(1):261. doi:10.1186/s12939-023-02072-5

7. Egger M, Vogelgesang L, Reitelbach J, Bergmann J, Müller F, Jahn K. Severe Post-COVID-19 Condition after Mild Infection: Physical and Mental Health Eight Months Post Infection: A Cross-Sectional Study. *Int J Environ Res Public Health*. 2024;21(1). doi:10.3390/ijerph21010021

8. Hastie CE, Lowe DJ, McAuley A, et al. True prevalence of long-COVID in a nationwide, population cohort study. *Nat Commun*. 2023;14(1):7892. doi:10.1038/s41467-023-43661-w

9. Shah AD, Subramanian A, Lewis J, et al. Long COVID symptoms and diagnosis in primary care: A cohort study using structured and unstructured data in The Health Improvement Network primary care database. *PLoS One*. 2023;18(9 September):1-21. doi:10.1371/journal.pone.0290583

10. Tan S, Pryor AJG, Melville GW, Fischer O, Hewitt L, Davis KJ. The lingering symptoms of post-COVID-19 condition (long-COVID): a prospective cohort study. *Intern Med J*. 2024;54(2):224-233. doi:10.1111/imj.16251

11. Zhang Y, Chinchilli VM, Ssentongo P, Ba DM. Association of Long COVID with mental health disorders: a retrospective cohort study using real-world data from the USA. *BMJ Open*. 2024;14(2):1-8. doi:10.1136/bmjopen-2023-079267

12. Sugiyama A, Miwata K, Kitahara Y, et al. Long COVID occurrence in COVID-19 survivors. *Sci Rep*. 2022;12(1):6039. doi:10.1038/S41598-022-10051-Z

13. Rubin R. As their numbers grow, COVID-19 "long haulers" stump experts. *JAMA*. 2020;324(14):1381-1383. https://jamanetwork.com/

14. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med*. 2021;27(4):626-631. doi:10.1038/s41591-021-01292-y

15. Sudre CH, Murray B, Varsavsky T, et al. Author Correction: Attributes and predictors of long COVID. *Nature Medicine 2021 27:6*. 2021;27(6):1113-1120. doi:10.1038/s41591-021-01361-2

16. Siff EJ, Aghagoli G, Marin BG, Tobin-Tyler E, Poitevien P. SARS-CoV-2 transmission: a sociological review. *Epidemiol Infect*. 2020;148:e242. doi:10.1017/S095026882000240X

17. Power T, Wilson D, Best O, et al. COVID-19 and Indigenous Peoples: An imperative for action. *J Clin Nurs*. 2020;29(15-16):2737. doi:10.1111/JOCN.15320

18. Vang ZM, Ng E. The impacts of COVID-19 on immigrants and the healthy immigrant effect: Reflections from Canada. *Prev Med (Baltim)*. 2023;171:107501. doi:10.1016/J.YPMED.2023.107501

19. O' Mahony L, Buwalda T, Blair M, et al. Impact of Long COVID on health and quality of life. *HRB Open Res*. 2022;5. doi:10.12688/HRBOPENRES.13516.1

20. Mahbub Hossain M, Das J, Rahman F, et al. Living with "long COVID": A systematic review and meta-synthesis of qualitative evidence. *PLoS One*. 2023;18(2). doi:10.1371/JOURNAL.PONE.0281884

21. Wolf S, Zechmeister-Koss I, Erdös J. Possible long COVID healthcare pathways: a scoping review. *BMC Health Serv Res*. 2022;22(1):1-11. doi:10.1186/S12913-022-08384-6/FIGURES/2

22. Baillie K, Davies HE, Keat SBK, et al. Complement dysregulation is a predictive and therapeutically amenable feature of long COVID. *medRxiv*. Published online 2023:2023.10.26.23297597. https://www.medrxiv.org/content/10.1101/2023.10.26.23297597v1

23. Cervia-Hasler C, Brüningk SC, Hoch T, et al. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. *Science*. 2024;383(6680):eadg7942. doi:10.1126/science.adg7942

24. Klein J, Wood J, Jaycox JR, et al. Distinguishing features of long COVID identified through immune profiling. *Nature*. 2023;623(7985):139-148. doi:10.1038/s41586-023-06651-y

25. Kolaczyk ED, Lee MM, Liu J, Parker MS. We Need a (Responsible!) Data Science Rapid Response Network. *Harv Data Sci Rev*. 2021;3(3.1). doi:10.1162/99608f92.2794e78d

26. Leonelli S. Data Science in Times of Pan(dem)ic. *Harv Data Sci Rev*. 2021;3(3.1):2021. doi:10.1162/99608f92.fbb1bdd6

27. Louissaint N. Bringing every tool to the COVID-19 fight – what we need now. *Harv Data Sci Rev*. 2021;3(1):2021. doi:10.1162/99608f92.70f8cb98

28. Chung MK, Caboni M, Strandwitz P, D'Onofrio A, Lewis K, Patel CJ. Systematic comparisons between Lyme disease and post-treatment Lyme disease syndrome in the U.S. with administrative claims data. *EBioMedicine*. 2023;90. doi:10.1016/J.EBIOM.2023.104524

29. Andersen KM, Khan FL, Park PW, et al. Using a data-driven approach to define post-COVID conditions in US electronic health record data. *PLoS One*. 2024;19(4):e0300570. doi:10.1371/JOURNAL.PONE.0300570

30. Hamad AF, Walld R, Lix LM, Urquia ML, Roos LL, Wall-Wieler E. Data resource profile: the Manitoba multigenerational cohort. *Int J Epidemiol*. 2022;51(3):e65-e72. doi:10.1093/ije/dyab195

31. Katz A, Enns J, Smith M, Burchill C, Turner K, Towns D. Population data centre profile: The Manitoba Centre for Health Policy. *Int J Popul Data Sci*. 2020;4(2). doi:10.23889/ijpds.v4i2.1131

32. Smith M, Roos LL, Burchill C, et al. Health services data: managing the data warehouse: 25 years of experience at the Manitoba Centre for Health Policy. In: *Data and Measures in Health Services Research*. Springer US; 2015:1-26. doi:10.1007/978-1-4899-7673-4_3-1

33. Roos L. A research registry uses, development, and accuracy. *J Clin Epidemiol*. 1999;52(1):39-47. doi:10.1016/S0895-4356(98)00126-7

34. Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. *Soc Sci Med*. 2004;58(11):2231-2241. doi:10.1016/j.socscimed.2003.08.008

35. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011;32(1):91-108. doi:10.1146/annurev-publhealth-031210-100700

36. SAS Institute. SAS Version 15.1. 2023;(9.4).

37. Michelen M, Manoharan L, Elkheir N, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health*. 2021;6(9):e005427. doi:10.1136/BMJGH-2021-005427

38. Hayes LD, Ingram J, Sculthorpe NF. More Than 100 Persistent Symptoms of SARS-CoV-2 (Long COVID): A Scoping Review. *Front Med (Lausanne)*. 2021;8:750378. doi:10.3389/FMED.2021.750378/BIBTEX

39. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nature Reviews Microbiology 2023 21:3*. 2023;21(3):133-146. doi:10.1038/s41579-022-00846-2

40. Silva Filho T, Song H, Perello-Nieto M, Santos-Rodriguez R, Kull M, Flach P. Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Mach Learn*. 2023;112(9):3211-3260. doi:10.1007/s10994-023-06336-7

41. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

42. Lankham I, Slaughter M. Simple and efficient bootstrap validation of predictive models using SAS/STAT® software. In: *Proceedings of the SAS Global Forum 2020 Conferenc*. SAS Institute Inc; 2020. https://support.sas.com/resources/papers/proceedings20/4647-2020.pdf

43. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017;2017:1-14. doi:10.1155/2017/3762651

44. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc*. 2018;113(521):390-400. doi:10.1080/01621459.2016.1260466

45. Zang C, Zhang Y, Xu J, et al. Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative. *Nature Communications 2023 14:1*. 2023;14(1):1-14. doi:10.1038/s41467-023-37653-z

46. Guilcher SJT, Bai YQ, Wodchis WP, Bronskill SE, Kuluski K. An interrupted time series study using administrative health data to examine the impact of the COVID-19 pandemic on alternate care level acute hospitalizations in Ontario, Canada. *Canadian Medical Association Open Access Journal*. 2023;11(4):E621-E629. doi:10.9778/CMAJO.20220086

47. Binka M, Klaver B, Cua G, et al. An elastic net regression model for identifying long COVID patients using health administrative data: a population-based study. *Open Forum Infect Dis*. 2022;9(12):ofac640. doi:10.1093/ofid/ofac640

48. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27(4):601. doi:10.1038/S41591-021-01283-Z

49. Perlis RH, Santillana M, Ognyanova K, et al. Prevalence and correlates of long COVID symptoms among US adults. *JAMA Netw Open*. 2022;5(10):E2238804. doi:10.1001/JAMANETWORKOPEN.2022.38804

50. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz J V. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. 2022;22(4):e102-e107. doi:10.1016/S1473-3099(21)00703-9

51. Chevinsky JR, Tao G, Lavery AM, et al. Late Conditions Diagnosed 1-4 Months Following an Initial Coronavirus Disease 2019 (COVID-19) Encounter: A Matched-Cohort Study Using Inpatient and Outpatient Administrative Data-United States, 1 March-30 June 2020. *Clin Infect Dis*. 2021;73(Suppl 1):S5-S16. doi:10.1093/CID/CIAB338

52. Kosowan L, Sanchez-Ramirez DC, Katz A. Understanding symptoms suggestive of long COVID syndrome and healthcare use among community-based populations in Manitoba, Canada: an observational cross-sectional survey. *BMJ Open*. 2024;14(1):e075301. doi:10.1136/BMJOPEN-2023-075301

53. Zhu Y, Mahale A, Peters K, et al. Using natural language processing on free-text clinical notes to identify patients with long-term COVID effects. In: *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. ACM; 2022:1-9. doi:10.1145/3535508.3545555

54. Wong MCS, Huang J, Wong YY, et al. Epidemiology, symptomatology, and risk factors for long COVID symptoms: population-based, multicenter study. *JMIR Public Health Surveill*. 2023;9:1-15. doi:10.2196/42315

55. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. 2019;7(2):e12239. doi:10.2196/12239

56. Luo J, Zhang J, Tang HT, et al. Prevalence and risk factors of long COVID 6–12 months after infection with the Omicron variant among nonhospitalized patients in Hong Kong. *J Med Virol*. 2023;95(6). doi:10.1002/jmv.28862

57. Marra AR, Sampaio VS, Ozahata MC, et al. Risk factors for long coronavirus disease 2019 (long COVID) among healthcare personnel, Brazil, 2020-2022. *Infect Control Hosp Epidemiol*. 2023;44(12):1972-1978. doi:10.1017/ice.2023.95

58. Martínez-Ayala MC, Proaños NJ, Cala-Duran J, et al. Factors associated with long COVID syndrome in a Colombian cohort. *Front Med (Lausanne)*. 2023;10(December):1-9. doi:10.3389/fmed.2023.1325616

59. Morello R, Mariani F, Mastrantoni L, et al. Risk factors for post-COVID-19 condition (Long Covid) in children: a prospective cohort study. *EClinicalMedicine*. 2023;59:101961. doi:10.1016/j.eclinm.2023.101961

60. Rayner DG, Wang E, Su C, et al. Risk factors for long COVID in children and adolescents: a systematic review and meta-analysis. *World Journal of Pediatrics*. 2023;20(2):133-142. doi:10.1007/s12519-023-00765-z

61. United Nations. *Indigenous Peoples and the COVID-19 Pandemic: Considerations*.; 2020. Accessed January 8, 2024. https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2020/04/COVID19_IP_considerations.pdf

62. Greenwood M, de Leeuw S, Lindsay N. Challenges in health equity for Indigenous peoples in Canada. *The Lancet*. 2018;391(10131):1645-1648. Accessed May 15, 2021. https://www-sciencedirect-com.uml.idm.oclc.org/science/article/pii/S0140673618301776

63. Curtice K, Choo E. *Indigenous Populations: Left behind in the COVID-19 Response*. Vol 395.; 2020. doi:10.1016/S0140-6736(20)31242-3

64. Nickel NC, Phillips-Beck W, Enns JE, et al. COVID-19 diagnostic testing and vaccinations among First Nations in Manitoba: A nations-based retrospective cohort study using linked administrative data, 2020–2021. *PLoS Med*. 2024;21(2):e1004348. doi:10.1371/JOURNAL.PMED.1004348

65. Romero-Ibarguengoitia ME, Rodríguez-Torres JF, Garza-Silva A, et al. Association of vaccine status, reinfections, and risk factors with Long COVID syndrome. *Sci Rep*. 2024;14(1):1-9. doi:10.1038/s41598-024-52925-4

66. Komaroff AL, Lipkin WI. ME/CFS and long COVID share similar symptoms and biological abnormalities: road map to the literature. *Front Med (Lausanne)*. 2023;10. doi:10.3389/fmed.2023.1187163

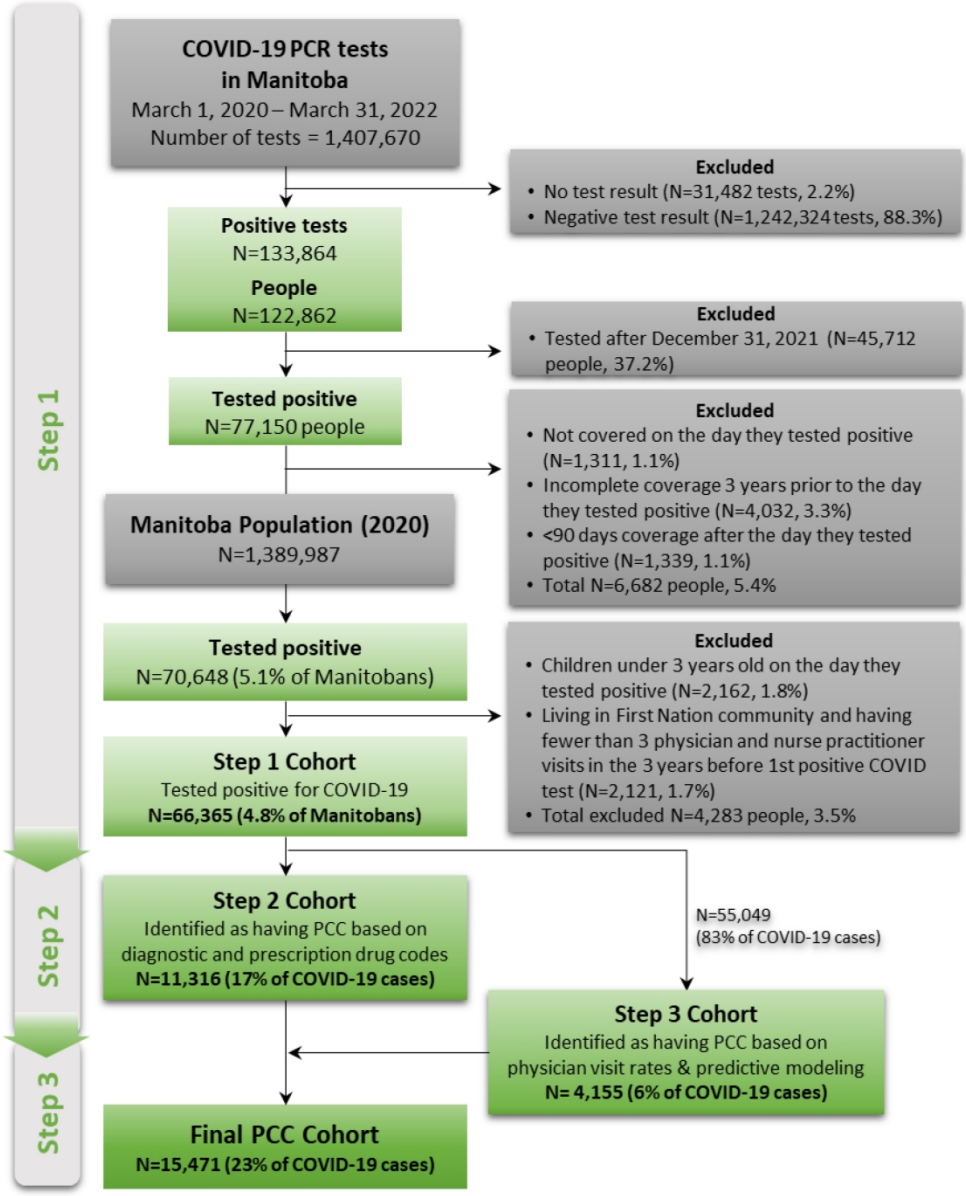


Figure 1. Post-COVID condition cohort development

148x182mm (300 x 300 DPI)

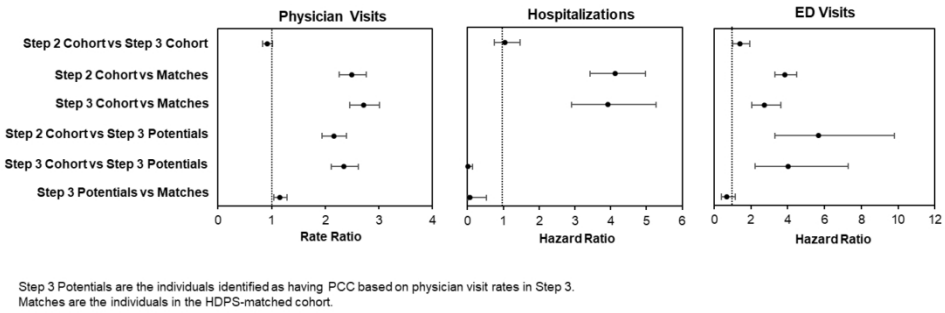


Figure 2. Three-way health service use comparison to examine validity of the physician visit/predictive modeling approach to identifying post-COVID condition (rate ratios and hazard ratios with 95% confidence intervals). ED: Emergency department.

256x99mm (300 x 300 DPI)

Supplemental Material

Table S1. ICD-9-CM codes, ICD-10-CA codes, and ATC codes used to identify post-COVID condition cases in Manitoba (Step 2 of cohort construction)

Disease Category	ICD-9-CM code (out-patient)	ICD-10-CA code (in-patient)	ATC code (prescriptions)
Endocrine, nutritional, and metabolic diseases		E10	
Mental, behavioral, and neurodevelopmental disorders	2962, 2968, 2969, 3004, 3007, 30742, 309, 311	F02, F03, F04, F07, F09, F32, F34, F39, F41, F43, F48, F90	N06A, N06BA
Diseases of the nervous system	3379, 3545, 3548, 3549, 3551, 3555, 3557, 3558, 3559, 3564, 3568, 3589, 3574, 3579	G40, G44, G471, G478, G479, G470, G50, G51, G52, G53, G55, G56, G57, G58, G59, G89, G904, G905, G908, G909, G933	N05B, N05C, N03AE
Diseases of the eye and adnexa		H103, H1042, H1040, H1050, H1089, H109, H162, H53	
Diseases of the ear and mastoid process	3861, 3883, 3887	H813, H814, H818, H819, H90, H920, H931, H932, H93A	
Diseases of the circulatory system	412, 4139, 4148, 4151, 4209, 4254, 4270, 4271, 4272, 4273, 4274, 4275, 4276, 4278, 4279, 428, 4290, 4293, 4298, 4299, 7850, 7851, 431, 432, 433, 434, 4359, 436, 452,4531, 4532, 4533, 4536, 4538, 4539, 4591,	I21, I2602, I30, I313, I314, I32, I40, I44, I45, I46, I47, I48, I49, I50, I5A	C07
Diseases of the respiratory system	4939	J310, J311, J312, J32, J370, J41, J42, J4521, J4522, J4531, J4532, J4541, J4542, J4551, J4552, J4590, J4599, J841	L04AX05, L01EX09, R03, H02A
Diseases of the digestive system		K293, K294, K296, K297, K298, K299, K290, K73, K744, K745, K7469, K7460, K740, K759	
Diseases of the skin and subcutaneous tissue		L25, L26, K299, L308, L309, L63, L65, L66, L984, L988, L989	
Diseases of the musculoskeletal system and connective tissue	7291	M028, M029, M048, M049, M128, M2480, M255, M259, M608, M609, M6281, M6283, M6289, M792, M797	M01

Diseases of the genitourinary system		N18, N394	
Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	7807, 7826	R002, R05, R0600. R0609, R0902, R093, R101, R103, R108, R109, R11, R198, R21, R32, R41, R42, R430, R49, R50, R51, R52, R538, R61, R630, R634, R7, R945,	N02
Injury, poisoning, and certain other consequences of external causes		T1491	
External causes of morbidity		X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83	

ATC: Anatomical Therapeutic Classification; ICD: International Classification of Disease

Table S2. Variables used in our predictive modeling approach to predict post-COVID condition among Manitobans who had tested positive for COVID-19 (Step 3 of cohort construction)

Variables considered in the training models	Variables included in the final model
Sex	**Y**
Age group at first positive COVID-19 test	**Y**
Region of residence (urban/rural)	
Regional Health Authority	**Y**
Income quintile	
SEFI-2 (socioeconomic factor index)	
Identified in medical records as having PCC in the first 90 days after positive COVID-19 test	**Y**
Hospital admission within 14 days of positive COVID-19 test	**Y**
Hospital admission after 14 days of positive COVID-19 test	**Y**
ICU admission within 14 days of positive COVID-19 test	**Y**
Charlson comorbidity index one year prior to positive COVID-19 test	**Y**
Number of COVID-19 vaccine doses received prior to positive COVID-19 test	**Y**
Recent immigrant to Manitoba (within five years of the study period)	
First Nations	

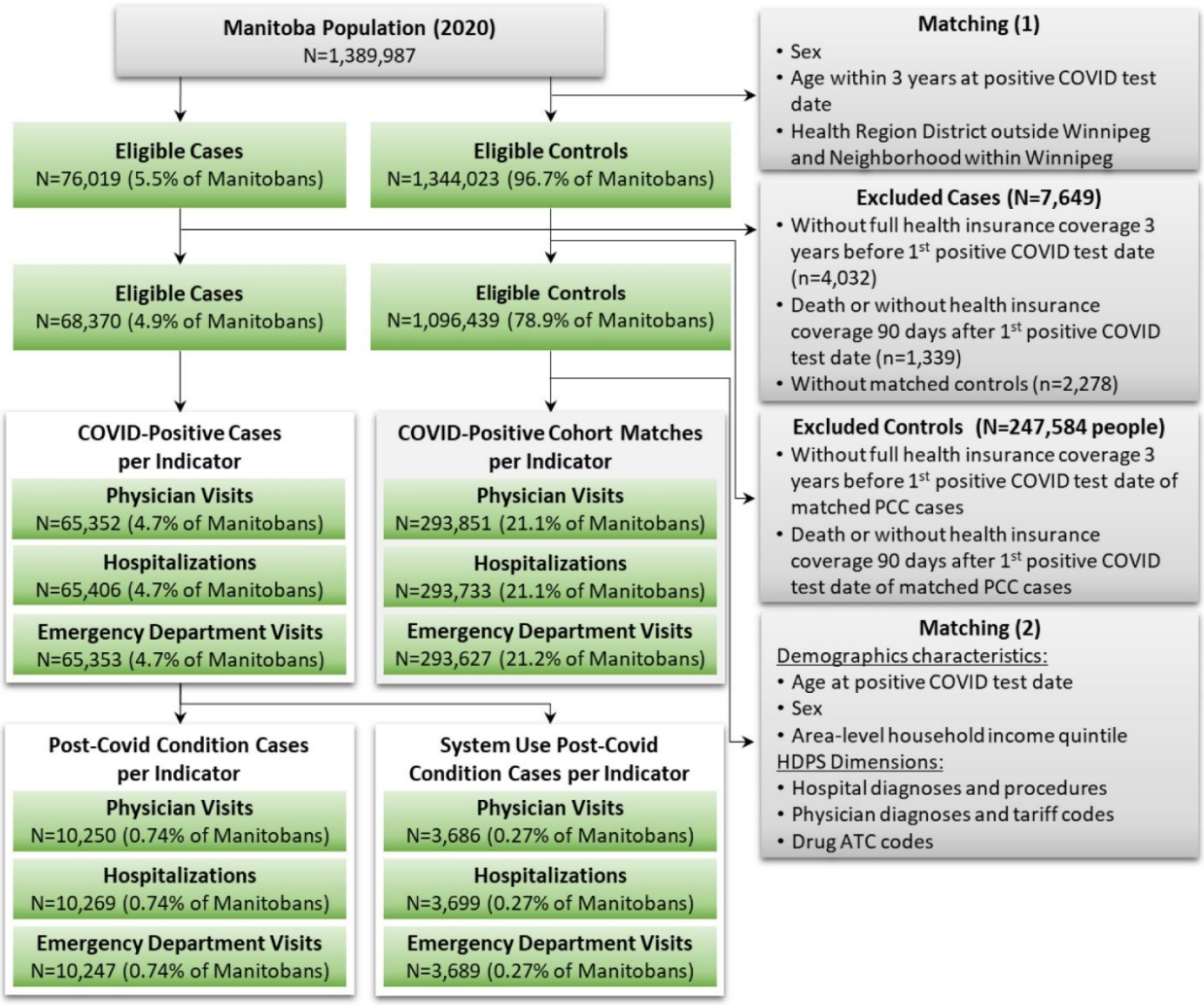


Figure S1. Development of the HDPS-matched cohorts for validation of Step 3 cohort construction