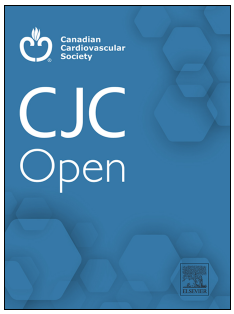
Journal Pre-proof



VALIDATION OF A CASE DEFINITION TO IDENTIFY PATIENTS DIAGNOSED WITH CARDIOVASCULAR DISEASE IN CANADIAN PRIMARY CARE PRACTICES

Riddhima Dinah Thomas, BSc, Leanne Kosowan, MSc, Mary Rabey, BMBS, Alan Bell, MD, Kim A. Connelly, MBBS PhD, Nathaniel M. Hawkins, MBChB, MD, MPH, Carolyn Gall Casey, BASc, Alexander G. Singer, MB BCh BAO

PII: S2589-790X(23)00089-6

DOI: https://doi.org/10.1016/j.cjco.2023.04.003

Reference: CJCO 662

To appear in: *CJC Open*

Received Date: 4 October 2022

Revised Date: 1 April 2023

Accepted Date: 17 April 2023

Please cite this article as: R.D. Thomas, L. Kosowan, M. Rabey, A. Bell, K.A Connelly, N.M Hawkins, C.G. Casey, A.G Singer, VALIDATION OF A CASE DEFINITION TO IDENTIFY PATIENTS DIAGNOSED WITH CARDIOVASCULAR DISEASE IN CANADIAN PRIMARY CARE PRACTICES, *CJC Open* (2023), doi: https://doi.org/10.1016/j.cjco.2023.04.003.

Cardiovascular Disease (CVD) in Canadian Primary Care Practices

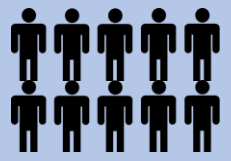
Retrospective Cross-sectional Study

Electronic Health Records (EMR) from 1,574 primary care providers

CPCSSN RCSSSP

Consenting to the Canadian Primary Care Sentinel Surveillance Network (CPCSSN)

Seven Canadian provinces



689,301 patients in the CPCSSN

2,484 patients in the CVD reference Standard

Validation of a CVD Case Definition

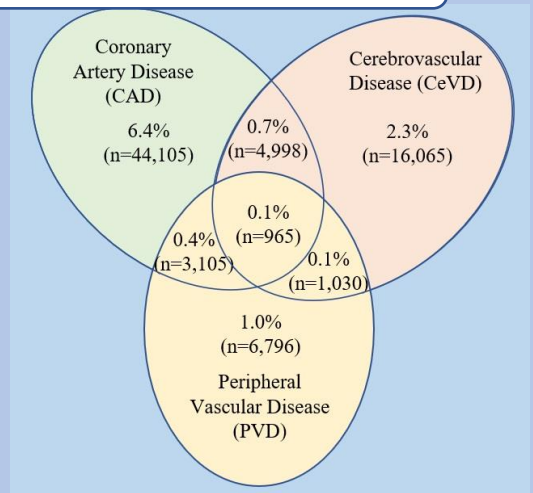
EMR

Case definition:
Billing records
Encounter diagnosis
Problem List

- ✓ Sensitivity 68.5%
- ✓ Specificity 97.8%
- ✓ PPV 77.7%
- ✓ NPV 96.5%

Prevalence of CVD in Primary Care

Estimated CVD prevalence in primary care 11.2%



Conclusion

1. The CVD case definition demonstrated expected prevalence and risk characteristics
2. Application can assess prevalence and disease burden for patients

VALIDATION OF A CASE DEFINITION TO IDENTIFY PATIENTS DIAGNOSED WITH CARDIOVASCULAR DISEASE IN CANADIAN PRIMARY CARE PRACTICES

Short Title: **CARDIOVASCULAR DISEASE IN CANADIAN PRIMARY CARE**

Riddhima Dinah Thomas, BSc, UCD School of Medicine, University College Dublin, Belfield, Dublin, Ireland

Leanne Kosowan, MSc, Department of Family Medicine, Rady Faculty of Health Sciences University of Manitoba, Winnipeg, Manitoba, Leanne.kosowan@umanitoba.ca

Mary Rabey, BMBS, Faculty of Medicine, University of Limerick, Limerick, Ireland

Alan Bell, MD, Department of Family and Community Medicine, University of Toronto, Toronto, Ontario

Kim A Connelly, MBBS PhD, Keenan Research Centre for Biomedical Science, Unity Health, St. Michael's Hospital, University of Toronto, Toronto, Ontario

Nathaniel M Hawkins, MBChB, MD, MPH, Centre for Cardiovascular Innovation, Division of Cardiology, University of British Columbia, Vancouver, British Columbia

Carolyn Gall Casey, BASc, Director Guidelines and Knowledge Translation, Canadian Cardiovascular Society

Alexander G Singer* MB BCh BAO, Department of Family Medicine, Rady Faculty of Health Sciences University of Manitoba, D09-780 Bannatyne Ave., Winnipeg, Manitoba, R3E 0W2, (204)272-3086, alexander.singer@umanitoba.ca (*corresponding)

Word count: 3,768/5000

Abstract

Background: Cardiovascular disease (CVD) is a leading cause of death globally. This study validates a primary care-based electronic medical record (EMR) case definition for CVD.

Methods: This retrospective cross-sectional study explores EMR data from 1,574 primary care providers participating in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). A reference standard was created by reviewing medical records of a subset of CPCSSN patients (n=2,017) for coronary artery disease (CAD), cerebrovascular disease (CeVD) and peripheral vascular disease (PVD). Together these produced a CVD reference. We applied validated case definitions to an active patient population (≥1 visit between 1-Jan-2018 and 31-Dec-2019) to estimate prevalence using exact binomial test (N=689,301). Descriptive statistics, chi-squared and t-tests characterized patients with and without CVD.

Results: The optimal CVD case definition 2 had a sensitivity 68.5% (61.6-74.8%), specificity 97.8% (97.0-98.4%), positive predictive value (PPV) 77.7% (71.6-82.7%), and negative predictive value (NPV) 96.5% (95.8-97.1%). Included in this CVD definition was a strong CAD case definition with sensitivity 91.6% (84.6-96.1%), specificity 98.3% (97.6-98.8%), PPV 74.8% (67.8-80.7%), and NPV 99.5% (99.1-99.7%). This CVD definition also included CeVD and PVD case definitions with low sensitivity (77.6% and 36.6%) but high specificity (98.6% and 99.0%). The estimated prevalence of CVD among primary care patients is 11.2% (CI 11.1-11.3) (n= 77,064), the majority had CAD (6.4%).

Conclusion: This study validated a definition of CVD and its component parts CAD, CeVD and PVD. Understanding the prevalence and disease burden for patients with CVD within primary care can improve prevention and disease management.

Introduction

Cardiovascular disease (CVD) refers to several conditions that affect blood vessels including coronary artery disease (CAD), cerebrovascular disease (CeVD) and peripheral vascular disease (PVD).[1-9] CVD contributes to significant direct and indirect healthcare costs. Globally CVD is the leading cause of mortality and accounts for approximately 30% of all deaths.[1,2,5] Most of the mortality associated with CVD relates to deaths from CAD and CeVD (i.e., strokes).[1,2,5-10] A third of these deaths are considered premature involving individuals under the age of 70 years.[1,2,7]

Underlying determinants of CVD are linked not only to hereditary factors but also to the social determinants of health[1,2]. Poverty, stress, urbanization, population ageing and globalization are all considered important drivers of CVD.[1] Specifically, patient risk factors for CVD include obesity, comorbidities, health behaviors, family history as well as psychosocial factors.[3-6,10-21] These risk factors often interact and amplify the risk of vascular damage leading to CVD morbidity and mortality at younger ages.[11,12] Canadian and international guidelines recommend broad implementation of measures in primary care settings to address CVD and related risk factors. [1,3-6,11,12] Identifying CVD within primary care can support improved prevention, treatment, and management, which in turn can prevent premature mortality.[1,3-6, 11,12]

Efforts have been made to utilize healthcare administrative databases for disease surveillance, however these efforts have not consistently been performed with primary care electronic medical record (EMR) data.[22] Although EMR-based case definitions

have been validated for several related conditions, CVD case definitions have largely been developed with administrative health data including hospital records.[23,24-34] CVD definitions based on administrative health data may lead to underreporting of some CVD presentations that precede hospitalization.[35-37] The widespread adoption of EMRs has resulted in the availability of primary care EMR data including information on diseases and their characteristics not currently captured in administrative data sources.[23,38] Primary care EMR data can improve case detection, thereby providing insight into the epidemiology of diseases and inform prevention, management and quality improvement strategies.[22,24,38-40] EMR-based primary care prevalence estimates can measure efforts to reduce disease burden through identification of patients and an understanding of primary and secondary prevention and treatments.

Clinical notes, diagnoses, examination results, laboratory data, risk factors, and prescribing practices are all important components of EMR data that have an enormous potential to be used to describe the care of conditions such as CVD.[23] We sought to develop, validate, and apply an EMR-based case definition of CVD with the goal of detecting and characterizing patients with CVD in primary care settings. Further, we characterize the prevalence and overlap between the component parts of CVD.

Materials and methods

This retrospective cross-sectional study assessed EMR data held in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). On a semi-annual basis, CPCSSN extracts and cleans de-identified longitudinal primary care EMR data from

seven Canadian provinces (British Columbia, Alberta, Manitoba, Ontario, Quebec, Nova Scotia, and Newfoundland and Labrador). Provincial networks extract EMR data from consenting family physicians, nurse practitioners and community pediatricians and transfer de-identified data to CPCSSN to create a single pan-Canadian EMR-based data repository representing >1,800,000 Canadians. Data are included for all patients that attended an appointment with a consenting provider unless they opt-out of inclusion. The repository includes structured data fields and short-text fields including billing, health condition(s) (problem list), encounter diagnoses, medication(s), exam, plus patient, and provider tables. A cohort was created of active patients (defined as those seen at least once in the past 2 years) from the CPCSSN repository with data available up to December 31, 2019. The starting point of data is dependent on when providers initiated their use of their EMR, with most records going back 10 or more years.

Reference set

Among active patients from the CPCSSN repository, we created a sub-set of randomly selected patients' EMR data for complete medical records review, to produce a positive and negative reference set for CVD. Random selection of records was performed using random number generator within SQL that accounted for the province of residency. Medical students and family medicine residents reviewed two files representing the EMR for 2,484 patients. The first was the clinical encounter note which contains a description of each encounter written by the provider and may include information on diagnosis, medication, blood pressures, and laboratory ordering. The second file included free-text entries in the health conditions, billing and encounter diagnosis table

representing the diagnosis name typed into the EMR by the primary care provider following the patient's encounter. Chart reviewers populated the data extraction table which included the following columns: diagnosis name, CVD (yes/no), type of CVD (i.e., CAD, CeVD, PVD), related conditions (e.g., MI, STEMI, stroke, vascular ulcer), plus documented symptoms and medications (Supplemental Appendix S2).

The Public Health Agency of Canada[7,8] suggests that the Canadian prevalence of heart disease is 8%. Using an estimated prevalence of at least 8%, a margin of error of 5%, and desired power of 80%, the sample size required is 201 positive cases. Our reference set has 203 positive cases and 1,814 negative cases (Figure 1). Patients within our positive reference set required a CVD diagnosis (CVD, CAD, CeVD, or PVD) be documented in the EMR by their primary care provider. Some providers did not include a specific CVD diagnosis and documented 'cardiovascular disease' as the diagnosis, whereas other providers included a specific diagnosis for CAD, CeVD, or PVD. Cases were further reviewed for the presence of symptoms, medications, and referrals to demonstrate the validity of the diagnosis. Records were reviewed by ≥2 medical students/residents. Discrepancies were highlighted and reviewed by a third reviewer. Negative cases did not include any indication of CVD in the EMR including diagnosis, medications, blood pressures, laboratory ordering or referrals.

Case definitions

We reviewed the literature for previously applied CVD case definitions[22,28-37] and tested our reference set compared to case definitions for CVD including separate definitions for CAD, CeVD, and PVD (Supplemental Appendix S1). Literature and clinical informed

case definitions produced 20 case definitions for CAD, 5 case definitions for CeVD, and 2 case definitions for PVD (Supplemental Appendix S1). CVD case definition 1 and 2 were created using CAD, CeVD, and PVD case definitions that had strong sensitivity, specificity, PPV and NPV (Supplemental Appendix S1). Case definition 15 and 18 were similar in their approach, we therefore chose to apply case definition 13 and 15 in CVD case definition 1 and 2, respectively. CVD case definitions 3 and 4 applied administrative-based CVD definitions (Table 1).[23,25-34] Our definition focused purely on the vascular aspects comprising CVD and omitted risk factors (i.e. hypertension) and other sequalae of cardiac disease (i.e. heart failure, arrhythmias and valvular diseases) that have been included in some administrative data based definitions to capture more general cardiac disorders. Patients captured with CAD, CeVD or PVD were considered a CVD positive case. We identified prescribed medications using ATC codes (cardiac therapy (ATC: C01DA02, C01DA05, C01DA08, C01DA14, C01EB09), beta-blockers (ATC: C07), calcium channel blockers (ATC: C08CA01, C08CA02, C08DA01), ACE inhibitors (ATC: C09A, C09B), and ARB (ATC: C09C, C09D)). We compared agreement between our reference set and CVD case definitions (Table 1).

Statistical Analyses

We compared the agreement of our reference set to case definitions using a 2x2 contingency table. We assessed the following metrics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and overall accuracy for CVD, CAD, CeVD, and PVD case definitions. The equations for these metrics are presented below.

$$PPV = \frac{TP}{}$$

$$\text{Sensitivity} = \frac{TP}{TP + FP}$$

$$\text{NPV} = \frac{TN}{TN + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

TP: true positive, TN: true negative, FN: false negative, FP: false positive

We used the exact binomial test to estimate prevalence and respective confidence intervals. We evaluated the CAD, CeVD, and PVD definitions using sensitivity, specificity, PPV and NPV to determine the best performing case definition. Using the best performing definition, we describe patients with and without CVD using descriptive statistics including mean, standard deviation (SD), and frequency. We explore differences in patients with and without CVD using Chi-square and t-tests. Patient age was calculated at the index date of December 31, 2019. We applied CPCSSN validated case definitions for hypertension and diabetes.[38] We identified conditions of interest using ICD-9 codes: dyslipidemia (ICD-9 code starting with 272), and heart failure (HF)[40] (ICD-9 codes starting with 428 or 425 or ACE/ARB and beta blocker and diuretic prescribed). We used the most recent body mass index (BMI) and blood pressure (BP)

recorded in the EMR for each patient. A BMI greater than or equal to 30 was considered obese. Visit frequency is an average of the mean number of visits in the previous three years (2017, 2018. 2019).

Statistical analyses were conducted using SAS V9.4 (SAS Institute Inc, Cary, NC).

This study was approved by the Health Research Ethics Board at the University of Manitoba. The research reported in this paper adhered to STROBE guidelines.

Results

Our reference set included 2,017 patients with 203 positive cases and 1,814 negative cases (Figure 1). Among the positive cases there were 107 patients with CAD, 67 patients with CeVD, and 41 patients with PVD (Figure 1). Patients may have been labelled as having multiple components of CVD. There were 15 patients with CAD and CeVD, 11 patients with CAD and PVD, <5 patients with CeVD and PVD, and <5 patients with CAD, CeVD, and PVD. There were 20 cases with documentation indicating 'cardiovascular disease' but without sufficient details to determine the type of CVD. There were 467 patients excluded where medical record review was inconclusive and reviewers could not be certain if the patient did or did not have CVD. Using kappa statistic, the interrater agreement was 96.7% (95.1-98.4%). Our reference set included patients from AB (11.5%), ON (25.7%), QC (0.6%), NL (0.2%), MB (58.2%), NS (0.1%), and BC (3.7%). Representation in each province was influenced by the number of providers participating in CPCSSN and availability of free-text encounter notes within the repository. Patients within the reference set had a mean age of 57 years (compared to 42 years in the Canadian Census) and were more likely to be female (57.7% vs. 50.7%).[41]

There were four case definitions for CVD. CVD case definitions 1 and 2 were developed based on case definitions validated to capture CAD, CeVD, and PVD. Case definition 2 had the strongest agreement with our reference set. CVD case definition 2 was built using the best preforming CAD, CeVD and PVD case definitions (i.e. case definition 15 for CAD, and case definition 1 for CeVD/PVD) and had a sensitivity of 76.9% (70.4-82.5%), specificity 97.2% (96.3-97.9%), PPV 75.4% (69.8-80.2%), and NPV 97.4% (96.7-97.8%). CVD case definition 1 was similar to case definition 2 but used CAD case definition 13 requiring ≥2 prescriptions for a related medication, which reduced the sensitivity to 68.5% (61.6-74.8%). (Table 2)

More specifically, among the 20 case definitions for CAD (Supplemental Appendix S1), three had strong agreement with our CAD reference set. Case definitions 15 for CAD (i.e. ≥1 ICD9 410-414 in billing/encounter diagnosis or health condition tables) had (sensitivity 91.6% (84.6-96.1%), specificity 98.3% (97.6-98.8%), PPV 74.8% (67.8-80.7%), NPV 99.5% (99.1-99.7%)). This was similar to case definition 13 (i.e. ≥1 ICD9 410-414 in billing/encounter diagnosis or health condition tables, and ≥2 CVD meds) (sensitivity 72.0% (62.5-80.2%), specificity 99.1% (98.6-99.5%)) and case definition 17 (i.e. ≥1 health condition for 410-414, OR ≥2 billing or encounter ICD9 410-414) (sensitivity 79.4% (70.5-86.6%), specificity 98.9% (98.3-99.3%)).

There were five definitions tested for capturing CeVD (Supplemental Appendix S1). Two had agreement statistics around 70%. Case definition 1 (i.e. ≥1 ICD9 430-438 in billing/encounter diagnosis or health condition tables) had a sensitivity 77.6% (65.8-86.9%), specificity 98.6% (97.9-99.0%), PPV 65.0% (55.7-73.3%), NPV 99.2% (98.8-99.5%). Case definition 5 (i.e. ≥1 Health condition for 430-438, OR ≥2 Billing or

encounter ICD9 430-438) had a sensitivity 64.2% (51.5-75.5%), specificity 98.0% (98.699.5%), PPV 70.5% (59.3-79.6%), NPV 98.8% (98.3-99.1%).

There were two case definitions tested for PVD, case definition 1 (i.e. ≥1 ICD-9 440.xx or 443, 443.9 in billing/encounter diagnosis or health condition tables) (sensitivity 36.6% (22.1-53.1%), specificity 99.0% (98.4-99.4%), PPV 42.9% (29.3-57.6%), NPV 98.7% (98.4-99.0%)) and case definition 2 (≥2 ICD9 440.xx or 443, 443.9) (sensitivity 12.2% (4.1-26.2%)), specificity 99.8%(99.4-99.9%), PPV 50.0% (23.1-76.9%), NPV 98.2% (98.0-98.4%)) (Supplemental Appendix S1).

CVD case definitions 3 and 4 were based on administrative case definitions, which aimed to capture heart conditions more broadly and were inclusive of related illness and risk factors such as hypertension, atrial fibrillation and HF. Due to these inclusion criteria, case definition 3 and 4 had a much lower PPV (21.0%, CI 19.9-22.1, and 23.3%, CI 20.8-25.9%, respectively) (Table 2). However, when we assessed the agreement of CVD case definitions 3 and 4 within a reference set of patients with CVD, hypertension, HF or non-valvular atrial fibrillation, our agreement markedly improved. Case definition 3 had a sensitivity 97.3% (95.7-98.5%), specificity 76.6% (74.3-78.8%), PPV 63.9% (61.6-66.0%), and NPV 98.6% (97.7-99.1%) (Supplemental Appendix S1).

There were 689,301 active patients in the CPCSSN dataset. When compared to the Canadian Census, CPCSSN patients were more likely to be female (56.3% vs. 50.7%) and older (52 years vs. 42 years). Upon applying case definition 2 to the active CPCSSN patient population (n=689,301), the estimated prevalence of CVD among patients seen in primary care settings is 11.2% (CI 11.1-11.3) (n= 77,064). When this is broken down into type of CVD there were 7.7% (n=53,173) patients with CAD, 3.4%

(n=23,058) patients with CeVD, and 1.7% (n=11,896) patients with PVD. Case definition 1 provided a similar prevalence estimate (9.6%, CI 9.5-9.6%). As expected, case definitions 3 and 4 provided a much higher prevalence estimate (51.0%, CI 50.9-51.2% and 31.1%, CI 31.0-31.2%, respectively). Case definitions 3 and 4 represents a population with a larger group of CVD related conditions. Figure 2 describes the proportion and type of CVD identified among active CPCSSN patients along with the prevalence of various combinations. The majority had CAD (6.4%), followed by CeVD (2.3%), PVD (1.0%) and both CAD and CeVD (0.7%) (Figure 2).

Table 3 characterized patients captured with CVD case definition 2 compared to patients not captured with CVD case definition 2. Patients captured with CVD case definition 2, were significantly more likely to be male (55.4% vs. 42.3%, p=<.001) and older (70.5 (SD14.4) vs. 50.0 (18.7), p=<.001). Patients captured with CVD case definition 2 were more likely to be diagnosed with hypertension (61.7% vs 24.3%, p=<.001), dyslipidemia (72.7% vs 34.6%, p=<.001), and/or HF (19.7% vs 2.3%, p=<.001). They were also more likely to be obese (8.4% vs. 5.9%, p=<.001). Seventy-five percent of patients captured with CVD case definition 2 were prescribed a CVD medication compared to 24.7% not captured (<.0001) (Table 3).

Discussion

The creation and validation of an EMR-based case definition for CVD will be beneficial for disease surveillance as well as research and quality improvement. CVD case definition 2 had strong sensitivities, specificities, PPV and NPV compared to our reference data reflecting satisfactory capture based on the available documentation. Our EMR-based CVD case definition 2 included CAD, CeVD, and PVD and suggests an

overall prevalence of 11.2% within primary care settings, with some patients managed in primary care diagnosed with more than one of these conditions (i.e., CAD, CeVD and PVD). Further, among the patients captured with CVD case definition 2, we found an association with several expected risk factors and comorbidities.[11-15]

Application of our CVD, CAD, CeVD and PVD definitions found a prevalence of 11.2% 7.7%, 3.7% and 1.7%, respectively. Related literature has defined CVD sub-types including acute myocardial infarction, ischemic heart disease (IHD) and CAD within similar ranges.[24-30] An analysis from the Global Burden of Disease Study suggests that 9% of the population is diagnosed with IHD, HF or stroke.[42]

Much of the previous literature on CVD case validation has defined individual types of CVD and often used administrative data that was not generalizable to a broad primary care population.[25-34] In fact, application of administrative case definitions in our study produced a prevalence of 51.0% and 31.1%, respectively. Few studies have examined CVD overall, which is relevant to understand the combined burden of atherosclerotic disease, which shares common risk factors and primary and secondary prevention strategies.[24,32,39] Additionally, there is variability in the existing literature regrading application of case definitions only to adult population or not requiring an age modifier. Our CVD case definition which included CAD, CeVD, and PVD reported a sensitivity of 76.9% and a PPV of 75.4%. The strong sensitivity of our case definition highlights its potential application for understanding CVD epidemiology.[43]

We expect that our CVD case definition was impacted by the performance of our case definitions for CAD, CeVD and PVD. In primary care CAD is the most frequently diagnosed of these CVD types. Case definition 15 had a sensitivity of 91.6% and a PPV of 74.8%. Tu et al. used a linked dataset of administrative data and EMR-billing records to validate a definition for ischemic heart disease, reporting a lower sensitivity of 77.0% and similar PPV of 78.8%.[26] The high sensitivity of the CAD definition can support studies aimed at inequalities in outcomes and epidemiology. The moderate PPV suggests the CAD definition could be used cautiously to define cohorts for medication studies.

CeVD case definition 1 did not perform as well as the CAD definition it had a sensitivity of 77.6% and a PPV of 65.0% and an estimated prevalence of 3.4%. Tu et al. validated a definition for stroke/transient ischemic attack using administrative data and reported a sensitivity of 68.0% and a prevalence of 3.0% which is similar to our study.[25] PVD had a low sensitivity of 36.6% and PPV of 42.9% and should be used with caution. PVD is underdiagnosed and undertreated in primary care settings despite development of diagnostic tools and management guidelines.[37] Peripheral vascular disease often goes undiagnosed because the physical findings vary widely, and it can be difficult to diagnose because the symptoms may be more subtle.[36] This PVD definition should be used in combination with CAD and CeVD definitions to capture patients with CVD.

Age and sex are important risk factors for CVD. Similar to previous literature, our case definition found a male predisposition for CVD and demonstrated strong associations

with age, obesity and hypertension.[3-6,44-46] Patients captured with our case definition 2 have similar characteristics to previous literature lending credence to the accuracy and utility of our definition. As the Canadian population ages, an improved understanding of the epidemiology of CVD and its relationship to CVD risk factors can better inform and prepare for healthcare system needs.[43,45] Application of a valid CVD case definition can inform disease epidemiology, which often highlights modifiable risk factors and has the potential to yield efficient utilization of time and resources by researchers and policy makers. Additionally, it can streamline the disease surveillance mechanisms by characterizing the population at risk.[43,47]

In most healthcare systems in high income countries such as Canada, primary care providers are often the stewards of healthcare as they are frequently the first point of contact for patients. Primary care providers are involved in primary prevention, diagnosis, treatment and secondary prevention of diseases for their patients. In such settings, having a tool that facilitates improved disease surveillance can inform management algorithms and policies that are designed to manage the ever-rising cases of CVD. [16-21] As demonstrated in our study, patients captured by CVD case definition 2 visited their primary care provider more frequently each year compared to patients not captured by CVD case definition 2. Visit frequency can be attributed to both CVD as well as multiple comorbidities and risk factors such as hypertension and diabetes. Evidence has demonstrated that primary and secondary prevention using evidence-based interventions can reduce CVD mortality and morbidity.[16-21] This merits attention for primary care system design and resourcing aimed at both primary and secondary prevention of CVD.[3-6,10-12] Nevertheless, addressing these risk factors with evidence-

based interventions in primary care is possible and likely to have a greater impact on those with CVD.[3-6,48-50] Since CVD impacts multiple organs (i.e. brain, heart, peripheral vasculature) it is co-managed by multiple practitioners, again re-enforcing the need for tools to support primary care providers who often frequently support care coordination and navigation for patients with multiple co-morbid conditions.[9]

Limitations

Whilst this study validated a case definition for CVD based on primary care EMR data with reasonable face validity, it is not without several limitations. The CPCSSN dataset includes a large number of patients and providers from various practice types and locations across Canada however, we are uncertain if our cohort is representative of other jurisdictions. Patients within the CPCSSN dataset were more likely to be female and older than the Canadian Census. The CPCSSN represents a population that had sought and accessed primary healthcare, future research should age- and sex-standardize prevalence estimates to describe prevalence estimates in a general Canadian population. This study accessed primary care EMR data for secondary purposes, but was unable to assess the accuracy of the documentation. This means that it is possible that in some instances, CVD or its component parts may be over- or under-diagnosed by the health care provider. This is less likely for serious conditions such as CAD and CeVD, which often have symptoms that are noticeable, require specific care in community settings and usually affect activities of daily living. It is possible that PVD diagnoses may have been omitted or underdiagnosed in some cases,[38-40] especially given the fact that it is known to be underdiagnosed and often managed outside of primary care, hence the less robust agreement.[36] A related issue is

that in some cases the medical records assessed do not specify the type of CVD, therefore using a broader CVD definition to describe patients may be more accurate when compared to capture of a type of CVD. Previous studies have demonstrated imperfect data quality of EMR records.[51,52] We cannot account for specialist consultations or specific investigations and operative reports (i.e., angiography, stress testing) that may be absent in primary care data, but would be present in administrative or hospital based data. However, our focus on diagnosis of CVD in primary care settings informs secondary prevention strategies provided by primary care providers. Future studies aiming to describe the complete spectrum of care could link primary and tertiary care data for CVD as well as assess unstructured data within a large and robust training set.

Conclusions

This study describes a validated highly specific definition of CVD, which is comprised of CAD, CeVD and PVD. The CVD cohort described has patient characteristics including comorbid conditions and risk factors consistent with the literature. Understanding the prevalence and disease burden for patients with CVD being managed in primary care settings can improve care for this important disease.

Acknowledgements

The authors would like to acknowledge William Peeler for his assistance in data curation.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Disclosures

Dr. Bell has received honoraria or grants from the following commercial and non-commercial organizations for consulting, speaking or research: Amgen, Bristol Myers Squibb, Janssen, AstraZeneca, Novartis, Pfizer, Bayer, Lilly, Boehringer Ingelheim, HLS Therapeutics, Sanofi, Bausch Health, Thrombosis Canada, Hypertension Canada, Canadian Cardiovascular Society. All other authors have no conflicts of interest to declare.

Data Statement

The datasets generated and/or analyzed during the current study are not publicly available due to the confidential nature of the data governed by PHIA legislation but are available from the corresponding author on reasonable request.

References

1 World Health Organization (2017) *Cardiovascular diseases.* Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (Accessed 1 June 2021).

2 Bansilal S, Castellano JM, Fuster V. Global burden of CVD: focus on secondary prevention of cardiovascular disease. International Journal of Cardiology. 2015;201(S1):S1-S7.

3 Graham I, Atar D, Corch-Johnsen K, Boysen G, Burell G, Cifkova R, et al. European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Caridology and Other Societies on Cardovascular Disease Prevention in Clinical Practice. (Constituted by representatives of nine societies and by intied experts). European Heart Journal. 2007;28(9):2375-2414. https://doi.org/10.1093/eurheartj/ehm316

4 Visseren FLJ, Mach R, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the task Force for Cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). European Hear Journal. 2021;42(34):3227-3337. https://doi.org/10.1093/eurheartj/ehab484

5 Pearson, G.J. *et al.* (2021) '2021 Canadian cardiovascular society guidelines for the management of dyslipidemia for the prevention of cardiovascular disease in the adult',

Canadian Journal of Cardiology. Available at: https://doi.org/10.1016/j.cjca.2021.03.016 (Accessed 1 June 2021)

6 Arnett DK, Blumentha RS, Albert MA, et al 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease. A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;140:e596-e646.

7 Public Health Agency of Canada. Heart Disease in Canada. 2017. Retrieved online from https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/diseases-conditions/heart-disease-fact-sheet/heart-disease-factsheet-eng.pdf (accessed March 23, 2023).

8 Manuel DG, Leung M, Nguyen K, Tanuseputro P, Johansen H. Burden of cardiovascular disease in Canada. Can J Cardiol 2003;19(9):997-1004.

9 Abramson BL, Al-Omran MA, Anand SS, Albalawi Z, Coutinho T, Mestral C, et al. Canadian Cardiovascular Society 2022 guidelines for peripheral arterial disease. Canadian Journal of Cardiology. 2022; 38(5): P560-P587. DOI: https://doi.org/10.1016/j/cjca.2022.02.029

10 Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors and therapeutics. J Cell Physiol. 2019;234(10):16812-16823. Doi: 10.1002/jcp.28350.

11 D'Agostino, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M. and Kannel, W.B. (2008) 'General cardiovascular risk profile for use in primary care: the Framingham heart study', *Circulation,* 117, pp. 743–753. Available at: https://doi.org/10.1161/CIRCULATIONAHA.107.699579 (Accessed 3 June 2021).

12 Cooney, M.T., Dudina, A., D'Agostino, R. and Graham, I.M. (2010) 'Cardiovascular Risk-Estimation Systems in Primary Prevention', *Circulation,* 122, pp. 300–310. Available at: https://doi.org/10.1161/CIRCULATIONAHA.109.852756 (Accessed 4 June 2021).

13 Akil, L., Ahmad, H.A., 2011. Relationships between Obesity and Cardiovascular Diseases in Four Southern States and Colorado. J Health Care Poor Underserved 22, 61–72. https://doi.org/10.1353/hpu.2011.0166

14 Wu, C.-Y., Hu, H.-Y., Chou, Y.-J., Huang, N., Chou, Y.-C., Li, C.-P., 2015. High Blood Pressure and All-Cause and Cardiovascular Disease Mortalities in Community-Dwelling Older Adults. Medicine (Baltimore) 94, e2160. https://doi.org/10.1097/MD.0000000000002160

15 Lee D, Chiu M, Manuel D, Tu Karen, Wang X, Austin P, et al. Trends in risk factors for cardiovascular disease in Canada: temporal, socio-demographic and geographic factors. CMAJ. 2009;181(3-4):E55-66. Doi: 10.1503/cmaj.081629

16 Brar R, Katz A, Ferguson T, Whitlock RH, Nella MD, Bohm C, Rigatto C, et al. Association of membership at a medical fitness facility with adverse health outcomes. American Journal of Preventative Medicine. 2021;61(5):e215-e224. https://doi.org/10.1016/j.amepre.2021.05.011

17 Barengo NC, Antikainen R, Borodulin K, Harald K, Jousilahti P. Leisure-time physical activity reduces total and cardiovascular mortality and cardiovascular disease incidence in older adults. J Am Geriatr Soc. 2017;65(3):504–510. https://doi.org/10.1111/jgs.14694.

18 Nocon M, Hiemann T, M€uller-Riemenschneider F, Thalau F, Roll S, Willich SN. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. Eur J Cardiovasc Prev Rehabil. 2008;15(3):239–246. https://doi.org/10.1097/HJR.0b013e3282f55e09.

19 O'Donovan G, Lee IM, Hamer M, Stamatakis E. Association of "weekend warrior" and other leisure time physical activity patterns with risks for all-cause, cardiovascular disease, and cancer mortality. JAMA Intern Med. 2017;177(3):335–342. https://doi.org/10.1001/jamainternmed.2016.8014.

20 Hamer M, Bauman A, Bell JA, Stamatakis E. Examining associations between physical activity and cardiovascular mortality using negative control outcomes. Int J Epidemiol. 2019;48(4):1161–1166. https://doi.org/10.1093/ije/dyy272.

21 Kosowan L, Shannon S, Rothney J, Halas G, Enns J, Holmqvist M, Wener P, Goertzen L, Katz A. Informing the physical activity evaluation framework: A scoping review of reviews. American Journal of Health Promotion. 2022;36(2):340-366. DOI: 10.1177/08901171211050059

22 Cozzolino, F., Abraha, I., Orso, M., Mengoni, A., Cerasa, M.F., Eusebi, P., Ambrosio, G. and Montedori, A. (2017) 'Protocol for validating cardiovascular and cerebrovascular ICD-9-CM codes in healthcare administrative databases: the Umbria Data Value Project', *BMJ Open,* 7. Available at: https://doi.org/10.1136/bmjopen-2016-013785 (Accessed 29 June 2021).

23 Wei, W.-Q., Teixeira, P.L., Mo, H., Cronin, R.M., Warner, J.L. and Denny, J.C. (2016) 'Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance', *Journal of the American Medical*

Informatics Association, 23, pp. 20–27. Available at: https://doi.org/10.1093/jamia/ocv130 (Accessed 29 June 2021).

24 McBrien, K.A., Souri, S., Symonds, N.E., Rouhi, A., Lethebe, B.C., Williamson, T.S., Garies, S., Birtwhistle, R., Quan, H., Fabreau, G.E. and Ronksley, P.E. (2018) 'Identification of validated case definitions for medical conditions used in primary care electronic medical record databases: a systematic review', Journal of the American Medical Informatics Association, 25, pp. 1567–1578. Available at: https://doi.org/10.1093/jamia/ocy094 (Accessed 29 June 2021).

25 Tu K, Wang M, Young J, Green D, Ivers N, Butt D, Jaakkimainen L, Kapral M. Validity of administrative data for identifying patients who have had a stroke or transient attach using EMERALD as a reference standard. Canadian Journal of Cardiology. 2013;29:1388-1394.

26 Tu K, Mitiku T, Lee D, Guo H, Tu J. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the electronic medical record administrative data linked database (EMRALD). Can J Cardiol. 2010;26(7):e225-e228.

27 Ammann EM, Schweizer ML, Robinson HG, Eschol JO, Kafa R, Girotra S, Winiecki S, et al. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for acute myocardial infraction (AMI) among intravenous immune globulin (IGIV) users in the sentinel distributed database. Pharmacoepidemiol Drug Saf. 2018;27(4):398-404. Doi:101002/pds.4398

28 Ahmad FS, Chan C, Rosenman MB, Post WS, Fort DG, Greenland P, Liu KJ, et al. Validity of Cardiovascular data from electronic sources: The multi-ethnic study of

antherosclerosis and HealthLNK. Circulation. 2017;136(13):1207-1216. Doi:10.1161/CIRCULATIONAHA.117.027436.

29 Goyal A, Norton CR, Thomas TN, Davis RL, Butler J, Ashok V, Zhao L, et al. Predictors of incident heart failure in a large insured population: A one million person-year follow-up study. Circ Heart Fail. 2010; 3:698-705. DOI: 10.1161/CIRCHEARTFAILURE.110.938175

30 Kivimaki M, Batty GD, Singh-Manoux A, Britton A, Brunner E, Shipley M. Validity of cardiovascular disease event ascertainment using linkage to UK hospital records. Epidemiology;2017:28(5):735-739. Doi: 10.1097/EDE.0000000000000688

31 Finlayson G, Ekuma O, Yogendran M, Burland E, Forget E. The Additional Cost of Chronic Disease in Manitoba. Winnipeg, MB: Manitoba Centre for Health Policy, 2010.

32 Lix L, Sobhan S, St-Jean A, Daigle JM, Fisher A, Yu OH. Et al. Validity of an algorithm to identify cardiovascular deaths from administrative health records: A multi-database population-based cohort study. BMC Health Services Research. 2021;21:758. https://doi.org/10.1186/s12913-021-06762-0

33 Katz A, Martens P, Chateau D, Bogdanovic B, Koseva I, McDougall C, Boriskewich E. Understanding the Health System Use of Ambulatory Care Patients. Winnipeg, MB: Manitoba Centre for Health Policy, 2013.

34 Chartier M, Dart A, Tangri N, Komenda P, Walld R, Bogdanovic B, Burchill C, Koseva I, McGowan K-L, Rajotte L. Care of Manitobans Living with Chronic Kidney Disease. Winnipeg, MB: Manitoba Centre for Health Policy, 2015.

35 Bhagirath VC, Nash D, Wan D, Anand SS. Building your peripheral artery disease toolkit: Medical management of peripheral artery disease in 2022. Canadian Journal of Cardiology. 2022; 38(5): 634-644. DOI:https://doi.org/10.1016/j.cjca.2022.02.004

36 Sontheimer D L. Peripheral Vascular Disease: Diagnosis and Treatment – American Family Physician [Internet]. 2006. [cited 2022 May 2]. Available from: https://www.aafp.org/afp/2006/0601/p1971.html

37 Lee DW, Cavender MA. Guidelines for peripheral vascular disease: where is the evidence? Circulation: Cardiovascular Interventions. 2019;12(1). https://doi.org/10.1161/CIRCINTERVENTIONS.118.007561

38 Williamson, T. *et al.* (2014) 'Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records', *Annals of Family Medicine.* Available at: https://www.annfammed.org/content/12/4/367 (Accessed 2 June 2021).

39 Institute of Medicine (US) Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing Countries (2010) Fuster V, Kelly BB, editors. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. Washington (DC): National Academies Press (US); 2010. 4, Measurement and Evaluation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK45691/ (Accessed 2 June 2021).

40 Vijh R, Wong S, Grandy M, Peterson S, Ezzat AM, Gibb AG, Hawkins NM. Identifying heart failure in patients with chronic obstructive lung disease through the Canadian Primary Care Sentinel Surveillance Network in British Columbia: a case derivation study. CMAJ Open. 2021;9(2):e376-383. DOI:10.9778/cmajo.20200183

41 Statistics Canada. (2022). Age, Sex at Birth and Gender Reference Guide: Census of Population, 2021. (No. 98-500-X issue 2021014). Available at https://publications.gc.ca/collections/collection_2022/statcan/98-500-x/98-500-x2021014-eng.pdf (accessed March 23, 2023).

42 Alam S, Lang JJ, Drucker AM, Gotay C, Kozloff N, Mate K, Patten SB, et al. Assessment of the burden of disease and injuries attributable to risk factors in Canada from 1990 to 2016: an analysis of the Global Burden of Disease Study. CMAJ Open. 2019;7(1): e140-e148. DOI:10.9778/cmajo.20180137

43 Fuster V, Kelly BB. Epidemiology of Cardiovascular Disease [Internet]. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. National Academies Press (US); 2010 [cited 2022 May 2]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK45688/

44 Bots, S.H., Peters, S.A.E., Woodward, M., 2017. Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. BMJ

45 Rodgers, J.L., Jones, J., Bolleddu, S.I., Vanthenapalli, S., Rodgers, L.E., Shah, K., Karia, K., Panguluri, S.K., 2019. Cardiovascular Risks Associated with Gender and Aging. J Cardiovasc Dev Dis 6, 19. https://doi.org/10.3390/jcdd6020019

46 Joseph P, Leong D, McKee M, Anand SS, Schwalm JD, Teo K, Mente A, Yusuf S. Reducing the Global burden of cardiovascular disease, Part 1: the epidemiology and risk factors. Circ Res. 2017;121(6):677-694. Doi: 10.1161/CIRCRESAHA.117.308903.

47 Dahlöf B. Cardiovascular Disease Risk Factors: Epidemiology and Risk Assessment. The American Journal of Cardiology. 2010 Jan 4;105(1, Supplement):3A-9A.

48 Kosowan L, Katz A, Halas G, LaBine L, Singer A. Using information technology to assess patient risk factors in primary care clinics: Pragmatic Evaluation. JMIR Formative Research. 2021;5(2):e24382. Doi: 10.2196/24382

49 Kosowan L, Katz A, Halas G, Singer A. Patient perspectives on tablet-based technology to collect risk factor information in primary care. BMC Family Practice. 2021;22:103. https://doi.org/10.1186/s12875-021-01443-7

50 Cykert S, Keyserling TC, Pignone M, DeWalt D, Weiner BJ, Trogdon JG, et al. A controlled trial of dissemintation and implementation of a cardiovascular risk reduction strategy in small primary care practices. Health Ser Res. 2020;55(6):944-953. Doi: 10.1111/1475-6773.13571

51 Singer A, Yakubovich S, Kroeker AL, et al. Data quality of electronic medical records in Manitoba: do problem lists accurately reflect chronic disease billing diagnoses? J Am Med Inform Assoc 2016; 23(6): 1107–1112.

52 Singer A, Kroeker AL, Yakubovich S, et al. Data quality in electronic medical records in Manitoba: Do problem lists reflect chronic disease as defined by prescriptions? Can Fam Physician 2017; 63(5): 382–389.

Tables

Table 1: Cardiovascular Disease Case Definitions for validation in the reference set	
Case	CAD: ≥1 Health Condition, billing or encounter dx for ICD-9 410-414

definition 1	AND ≥2 medications for ATC codes starting with C01, C07, C08, C09 **OR** CeVD: ≥1 Health Condition, billing or encounter dx for ICD-9 430-438 **OR** PVD: ≥1 Health Condition, billing or encounter dx for ICD-9 440.xx, or 443.xx
Case definition 2	CAD: ≥1 Health Condition, billing or encounter dx for ICD-9 410-414 **OR** CeVD: ≥1 Health Condition, billing or encounter dx for ICD-9 430-438 **OR** PVD: ≥1 Health Condition, billing or encounter dx for ICD-9 440.xx, or 443.xx
Case definition 3	≥1 Health Condition, billing or encounter dx for ICD-9 390-429, 430-448,458 **OR** ≥1 ATC code from medication table for B01A, C01A, C01B, C01CA17, C01D, C02AA, C02AB, C02C, C02D, C02L, C03, C04AD, C05BA, C07AA01, C07AA02 C07AA03, C07AA04, C07AA06, C07AA07, C07AA12, C07AB, C07AG, C07B, C07C, C08, C09, C10
Case definition 4	≥1 Health Condition, billing or encounter dx within 1 year for ICD-9 390-429, 430-448,458 **OR** ≥1 ATC code from medication table within 1 year for B01A, C01A,

	C01B, C01CA17, C01D, C02AA, C02AB, C02C, C02D, C02L, C03, C04AD, C05BA, C07AA01, C07AA02, C07AA03, C07AA04, C07AA06, C07AA07, C07AA12, C07AB, C07AG, C07B, C07C, C08, C09, C10

CAD: coronary artery disease

CeVD: cerebrovascular disease

PVD: peripheral vascular disease

ICD-9: International Classification of Diseases, Ninth edition, Clinical Modification

ATC: Anatomical Therapeutic Chemical code

Table 2: Agreement between National Reference set for CVD and CVD case definitions*

N=2017

	Sen	Spec	PPV	NPV	Accuracy
Case definition 1	68.47 (61.6-74.8)	97.79 (97.01-98.42)	77.65 (71.61-82.72)	96.52 (95.77-97.14)	94.84 (93.79-95.77)
Case definition 2	76.85 (70.43-82.46)	97.19 (96.32-97.9)	75.36 (69.79-80.2)	97.4 (96.69-97.97)	95.14 (94.11-96.04)
Case definition 3	94.58 (90.51-97.26)	60.09 (57.79-62.35)	20.96 (19.9-22.06)	99.0 (98.24-99.44)	63.56 (61.42-65.66)
Case definition 4	59.11 (52.01-65.94)	78.17 (76.2-80.05)	23.26 (20.79-25.92)	94.47 (93.53-95.28)	76.25 (74.33-78.09)

* Supplemental Appendix S1includes the agreement for the CAD, CeVD and PVD definitions

Table 3: Patients with at least 1 visit to a primary care provider participating in CPCSSN between January 1, 2017 and December 31, 2019. N= 689,301			
Variable name	Patients without CVD n= 612,237	Patients with CVD (case definition 2) n= 77,064	P-Value
Patient sex (% male)	258,657 (42.3%)	42,644 (55.4%)	<.0001
Patient age (mean, (SD))	50.0 (18.7)	70.5 (14.4)	<.0001
Annual visit frequency (mean, (SD))	2.6 (3.0)	4.8 (5.5)	<.0001
Systolic blood pressure (mean, (SD))	125 (16)	130 (17)	<.0001
Diastolic blood pressure (mean, (SD))	76.8 (10)	74 (11)	<.0001
Obesity (BMI >=30)	35,965 (5.9%)	6,483 (8.4%)	<.0001
Diabetes	68,515 (11.2%)	23,169 (30.1%)	<.0001
Hypertension	148,654 (24.3%)	47,509 (61.7%)	<.0001
Heart Failure	14,003 (2.3%)	15,187 (19.7%)	<.0001

Dyslipidemia	211,529 (34.6%)	55,992 (72.7%)	<.0001
CVD medication	151,236 (24.7%)	57,924 (75.2%)	<.0001
Cardiac therapy	7,935 (1.3%)	18,917 (24.6%)	<.0001
Beta-blockers	44,886 (7.3%)	33,555 (43.5%)	<.0001
Calcium channel blockers	59,365 (9.7%)	25,720 (33.4%)	<.0001
ACE inhibitors	81,693 (13.3%)	37,563 (48.7%)	<.0001
ARB	47,782 (7.8%)	17,544 (22.8%)	<.0001

Figure Legend

Figure 1: Flow diagram for creation of the cardiovascular disease (CVD) reference set from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN).

Figure 2: Venn diagram demonstrating proportion of active patients in CPCSSN captured using the strongest CVD case definitions and the overlap of CAD, CeVD and PVD diagnosis in primary care settings.

Figure 1: Flow Diagram

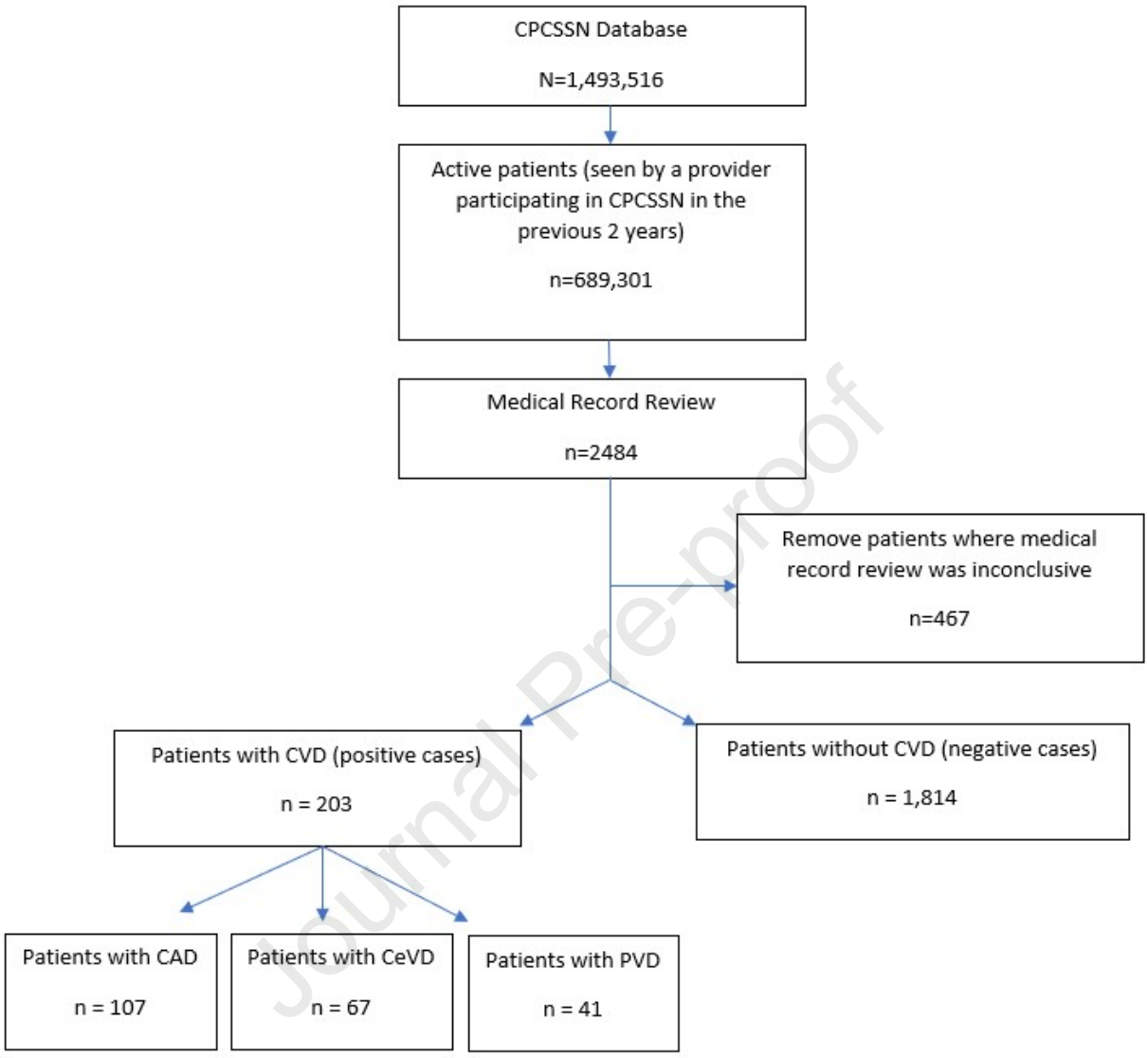


Figure 2 Venn Diagram
n=689,301