CJC Open 5 (2023) 709–712

Emerging Evidence

The Challenges of Identifying Patients With Peripheral Artery Disease Utilizing Administrative Databases

Pishoy Gouda, MB, BCh, BAO, MSc,[a] Douglas C. Dover, PhD,[b] Elinor (Mengxiao) Wang, MSc,[b] Victor Aboyans, MD, PhD,[c,d] and Robert Welsh, MD, MSc[a,b]

[a] *Mazankowski Alberta Heart Institute, University of Alberta, Division of Cardiology, Edmonton, Alberta, Canada*

[b] *Canadian VIGOUR Centre, University of Alberta, Edmonton, Alberta, Canada*

[c] *Department of Cardiology, Dupuytren-2 University Hospital, Limoges, France*

[d] *EpiMaCT, Inserm 1094 & IRD 270, Limoges University, Limoges, France*

ABSTRACT

Peripheral artery disease (PAD) carries a high burden of morbidity when identified in patients with coronary artery disease (CAD). However, identification of patients with concomitant CAD and PAD remains challenging. Using linked administrative databases of 207,026 individuals with CAD between 2002 and 2019 (median follow-up, 4.7 years), a model for PAD was applied to identify baseline PAD and the development of PAD during follow-up. Both baseline PAD and future PAD models demonstrated poor calibration and discrimination (c-statistic 0.618 and 0.583). In the absence of additional variables, the present models are unable to identify patients with concomitant CAD and PAD.

RÉSUMÉ

La maladie artérielle périphérique (MAP) impose un lourd fardeau de morbidité lorsqu'elle est diagnostiquée chez les patients atteints de coronaropathie. Toutefois, il reste difficile de repérer les patients atteints à la fois de coronaropathie et de MAP. À partir de bases de données administratives liées comptant 207 026 personnes atteintes de coronaropathie entre 2002 et 2019 (suivi médian de 4,7 ans), un modèle pour la MAP a été appliqué afin de repérer une MAP initiale et l'apparition d'une MAP au cours du suivi. Les modèles de MAP initiale et de MAP future ont tous deux été associés à un calibrage et à une capacité de distinction insatisfaisants (statistique C de 0,618 et 0,583). En l'absence d'autres variables, les modèles actuels sont incapables de repérer les patients atteints de coronaropathie et de MAP concomitantes.

The prevalence of peripheral artery disease (PAD) is estimated to be 5.6%, affecting ∼236 million individuals around the world and ∼2 million Canadians.[1] The presence of even asymptomatic PAD is associated with a 3-fold increase in major cardiovascular events,[2] which has led to renewed efforts to identify these patients earlier in their disease trajectory. Such identification remains challenging, due to lack of awareness on the part of the public, and the fact that up to 50% of patients with PAD are asymptomatic.[3] Although the ankle-brachial index (ABI) remains the noninvasive gold standard for diagnosing PAD, screening for asymptomatic PAD has not been recommended, owing to a lack of evidence to suggest that it improves outcomes or is cost-effective. Compared to the general population, patients with coronary artery disease (CAD) are twice as likely to have PAD.[4] In addition, in the **C**ardiovascular **O**utcom**e**s for **P**eople Using **A**nticoagulation **S**trategie**s** (COMPASS) trial, patients with both CAD and PAD demonstrated the greatest benefit from the addition of low-dose rivaroxaban therapy to aspirin to reduce the incidence of major adverse cardiovascular events (MACEs).[5] In the absence of a routine ABI screening program, alternative methods to identify patients with CAD who have concomitant asymptomatic PAD are required. One approach is to harness administrative data to provide contemporary data on the prevalence and outcomes of this high-risk population.

As administrative databases are reliant on coding of physician claims, the fact that previous studies have demonstrated that International Classification of Diseases (ICD) coding has poor diagnostic characteristics and fails to identify asymptomatic PAD should not be surprising.[6] Using an adapted version of a diagnostic model derived from participants in the COMPASS trial,[7] we aimed to ascertain its usefulness in an administrative database of patients with CAD.

Received for publication April 4, 2023. Accepted June 14, 2023.

Corresponding author: Dr Robert C. Welsh, Mazankowski Alberta Heart Institute, 2C2 Cardiology, WMC 8440 112 St NW, Edmonton, Alberta T6G 2B7, Canada. Tel.: +1-780-407-3613; fax: +1-780-407-3639.

E-mail: Robert.Welsh@ahs.ca

See page 712 for disclosure information.

https://doi.org/10.1016/j.cjco.2023.06.003

Table 1. Patient demographics stratified by presence of baseline peripheral arterial disease (PAD)

Characterisic	No PAD at baseline	PAD at baseline	Overall	*P*
Total, N (%)	198,454 (95.9)	8572 (4.1)	207,026 (100.0)	
Age, y				
Mean (SD)	66.9 (13.0)	71.6 (11.5)	67.1 (12.9)	< 0.0001
Median (IQR)	67.0 (58.0, 77.0)	72.0 (64.0 80.0)	67.0 (58.0, 77.0)	< 0.0001
Age group				
< 60	56,815 (28.6)	1310 (15.3)	58,125 (28.1)	< 0.0001
60−69	56,460 (28.4)	2179 (25.4)	58,639 (28.3)	—
70−79	48,294 (24.3)	2760 (32.2)	51,054 (24.7)	—
≥ 80	36,885 (18.6)	2323 (27.1)	39,208 (18.9)	—
Sex, female	71,206 (35.9)	2982 (34.8)	74,188 (35.8)	0.0389
Serum creatinine, mg/dL, mean (SD)	92.7 (57.2)	135.1 (137.2)	94.4 (63.1)	< 0.0001
Cholesterol, umol/L, mean (SD)	4.4 (1.2)	3.9 (1.2)	4.3 (1.2)	< 0.0001
MI	62,134 (31.3)	3657 (42.7)	65,791 (31.8)	< 0.0001
Diabetes	56,460 (28.4)	4797 (56.0)	61,257 (29.6)	< 0.0001
Hypertension	137,082 (69.1)	7647 (89.2)	144,729 (69.9)	< 0.0001
TIA	6895 (3.5)	664 (7.7)	7559 (3.7)	< 0.0001
Stroke	11,574 (5.8)	1323 (15.4)	12,897 (6.2)	< 0.0001
Develop PAD	7371 (3.7)	—	7371 (3.6)	

Values are n (%), unless otherwise indicated.

IQR, interquartile range; MI, myocardial infarction; SD, standard deviation; TIA, transient ischemic attack.

Methods

Patient population

Alberta, with a population of approximately 4.5 million people, has a single-payer, publicly funded healthcare system. Data for the study were obtained by linking the following administrative healthcare databases using the unique provincial personal health number: the Canadian Institute for Health Information Discharge Abstract Database of all admissions to acute care facilities, containing a most-responsible diagnosis, and up to 24 other diagnoses and up to 20 procedure codes; the Canadian Institute for Health Information National Ambulatory Care Reporting System database of emergency visits and day procedures, containing 10 diagnosis fields and 10 procedure code fields; practitioner claims of physician billings to the province, containing up to 3 diagnosis codes; the Alberta Health Care Insurance Plan registry, tracking Albertans eligible for healthare in the province, used to determine residency and age; Alberta Vital Statistics Office—capturing the date of death; and laboratory data indicating the test and test results. Using these linked administrative healthcare databases, we identified a retrospective cohort of adult patients with CAD in Alberta, Canada between 2002 and 2019, with lab values available for both serum creatinine and cholesterol (within 6 months of each other, data available starting in 2012) defining the baseline date. The diagnoses of both CAD and PAD were determined by identifying relevant ICD codes from hospitalization records, ambulatory care visits, and practitioner claims. We required 2 identified instances in practitioner claims, at least 30 days apart within 1 year, or 1 instance from hospitalization or ambulatory records. Codes for CAD included the following: ICD-9: 410-414; and ICD-10: I70, I710-719, I731, I738, I739, I771, I790, I792, K551, K558, K559, Z958, and Z959. Codes for PAD included the following: ICD-9: 093.0, 437.3, 440, 441, 443, 447.1, 557.1, 557.9, and V43.4; and ICD-10: I70 I710-I712, I715-I719, I731, I738, I739, I771, I790, I792, K551, K558, K559, Z958, and Z959). The primary outcome was a composite MACE endpoint that included hospitalization for stroke or myocardial infarction, coronary revascularization, and death.

Table 2. Calibration and discrimination for reduced model 1 predicting symptomatic peripheral arterial disease

	Ideal (well calibrated)	External validation sample	*P*	c-statistic
Calibration-in-the-large*	—	−5.22 (intercept)	—	0.618
Calibration slope†	1	0.39 (slope)	< 0.0001	
Fully re-estimated	n/a	n/a	< 0.0001	0.743

n/a, not applicable .

* Intercept not reported, so no formal test is possible.

† Testing slope equal to 1.

Diagnostic models

The models utilized in the study have been described previously, as developed from participants in the COMPASS trial.[7] Derived from the COMPASS trial, symptomatic PAD was defined as known history of PAD, history of claudication, previous arterial bypass, previous amputation, or PAD identified on imaging studies. Asymptomatic PAD was defined as an ABI of < 0.9 in the absence of symptoms of PAD. The model components of body mass index, systolic and diastolic blood pressure, heart rate, and tobacco use were not available in the administrative health data and were omitted. The following variables were included: age, sex, serum creatinine, total cholesterol, history of myocardial infarction, diabetes, hypertension, transient ischemic attack, and stroke.

Statistical analysis

Model 1 was a logistic regression model predicting symptomatic lower-extremity PAD and carotid disease. so it was validated against PAD status at baseline according to administrative health records. Model 2 was a logistic regression model predicting asymptomatic lower-extremity PAD in those

Table 3. Calibration and discrimination for reduced model 2 predicting future peripheral arterial disease

	Binary outcome models				Time-to-event models
	Ideal (well calibrated)	External validation sample	*P*	c-statistic	Integrated AUC*
Calibration-in-the-large†	—	−10.4 (intercept)	—	0.583	—
Calibration slope‡	1	0.057 (slope)	< 0.0001		0.592
Fully re-estimated	n/a	n/a	< 0.0001	0.618	0.428

n/a, not applicable.

* For both models, the area under the receiver operating characteristic curve (AUC) was virtually constant over time. The time-dependent AUC in the calibration slope model ranged from 0.587 to 0.603, and in the fully re-estimated model, it ranged from 0.422 to 0.432.

† Intercept not reported, so no formal test is possible.

‡ Testing slope equal to 1.

without a PAD diagnosis at baseline, so it was validated against future diagnoses of PAD in the administrative data.

In model 1, calibration-in-the-large—the assessment of observed to model-predicted outcomes—was assessed visually using a calibration plot and by allowing the intercept to vary, holding the predictor effects fixed, in a logistic regression on the outcome PAD at baseline. Slope calibration—the assessment of systematic overestimation or underestimation—in model 1 was performed using a calibration plot and testing the slope coefficient equal to 1 in a logistic regression. A fully recalibrated model was assessed using a calibration plot and a logistic regression model including the original risk score as a fixed offset and allowing coefficients for all measures to vary, capturing the change in each coefficient. Discrimination in these models was assessed using the c-statistic.

Calibration-in-the-large, slope Calibration, and fully recalibrated models in model 2 were assessed as above, with the outcome of future PAD diagnosis. To fully account for time, cumulative event rates, shown as Kaplan-Meier curves of PAD incidence by decile of risk score, are presented and tested using the log-rank test. Discrimination was assessed using the c-statistic, and the time varying and integrated area-under-the-receiver-operating-characteristic-curve statistic of Uno.[8] Models assessing future PAD exclude patients who had PAD at baseline or who died on the index date.

Predictive validation evaluates the ability of a risk score to predict clinically relevant events beyond the original purpose of PAD prediction. The composite outcome of myocardial infarction, stroke, and death was predicted using both models. Kaplan-Meier curves, stratifying the predicted values from the slope-calibrated models into deciles, were tested using the log-rank test.

This study was approved by the University of Alberta Research Ethics Board (Pro00082215). The ethics panel determined that the research is a retrospective database review for which subject consent for access to personally identifiable health information would not be reasonable, feasible, or practical. All statistical analyses were carried out using SAS version 9.4 (SAS Institute, Cary, NC).

Results

We identified 207,026 individuals with CAD between 2002 and 2019, with both lab values present (serum

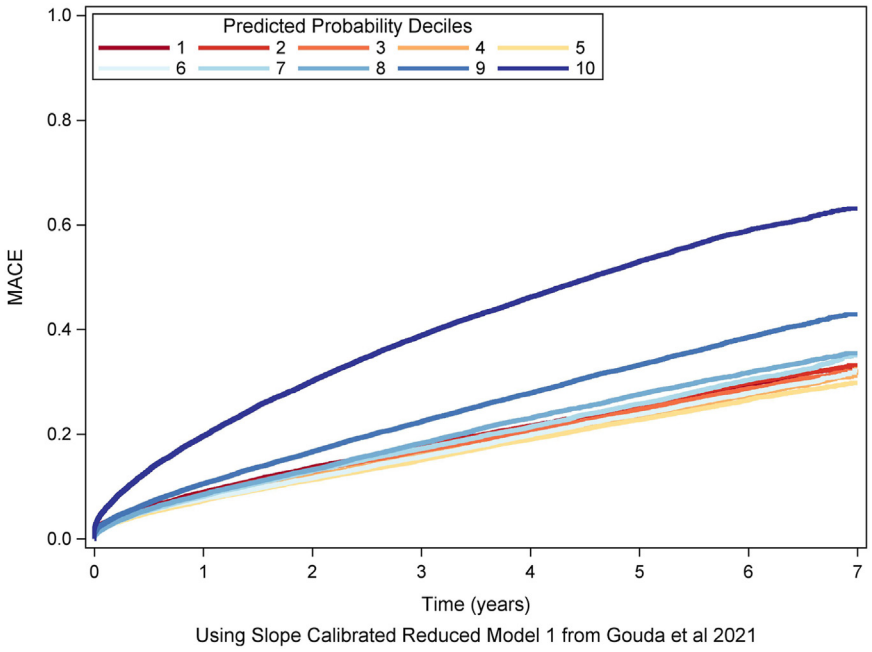

Figure 1. Kaplan-Meier curve of major adverse cardiovascular events (MACEs—composite of death, myocardial infarction, and stroke), stratified by deciles of the slope-calibrated model.

creatinine and total cholesterol). At baseline, 8572 (4.1%) had a diagnosis of PAD, and 7371 (3.6%) developed PAD over a median follow-up period of 4.7 years. The mean age of patients was 67.1 +/- 12.0 years, and 35.8% were female (Table 1).

Model 1 for the presence of PAD at baseline had poor calibration (rejected a calibrated slope, $P < 0.0001$) and had poor discrimination (c-statistic, 0.618; Table 2). After fully re-estimating the model coefficients (Supplemental Table S1), the model displayed good calibration and fair discrimination (c-statistic, 0.743). Model 2 for the development of future PAD demonstrated poor calibration (rejected a calibrated slope, $P < 0.0001$) and poor discrimination (c-statistic, 0.583; Table 3). After fully re-estimating the model coefficients (Supplemental Table S2), the model displayed good calibration and poor discrimination (c-statistic, 0.618).

Overall, the composite outcome occurred in 9.6%, 19.6%, and 37.0% of patients at 1, 3, and 7 years of follow-up. However, deciles of the PAD risk score were poor at discriminating between MACE risk with only the 2 largest risk-score quintiles separating from the rest (Fig. 1).

Conclusion

The models evaluated in the present study utilized variables commonly available in administrative databases. These models were not able to identify patients with PAD or those at risk of developing PAD, limiting the current capacity to utilize these databases for active patient identification. Further work is required to understand why PAD is frequently underdiagnosed, and how this situation can be improved in the clinical setting. One potential avenue lies in the introduction of simple-to-use ABI measuring devices that can be used easily at the bedside by frontline healthcare workers. The introduction of these devices into the market opens the possibility for the quick and reliable diagnosis of PAD. Whether advances in machine-learning algorithms can leverage structured and unstructured data in electronic medical records to accurately identify this underdiagnosed, undertreated, and high-risk population remains to be seen.

Acknowledgements

This study is based in part on data provided by Alberta Health and Alberta Health Services. The authors thank the Customer Relationship Management and Data Access Unit at Alberta Health for creating the linked database. The interpretation and conclusions are those of the researchers and do not represent the views of the Government of Alberta. Neither the Government of Alberta nor Alberta Health expresses any opinion in relation to this study.

Data Statement

All authors take responsibility for all aspects of the reliability and freedom-from-bias of the data presented and their discussed interpretation.

Ethics Statement

This study was approved by the University of Alberta Research Ethics Board (Pro00082215). The ethics panel determined that the research is a retrospective database review for which subject consent for access to personally identifiable health information would not be reasonable, feasible, or practical.

Patient Consent

The authors confirm that patient consent is not applicable to this article. This is a retrospective case report using de-identified data; therefore the IRB did not require consent from the patient.

Funding Sources

This work was partially funded by the CCS-Bayer Cardiovascular Research Award, awarded to P.G.

Disclosures

The authors have no conflicts of interest to disclose.

References

1. Song P, Fang Z, Wang H, et al. Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study. Lancet Glob Health 2020;8:e721-9.
2. Ankle Brachial Index Collaboration, Fowkes FGR, Murray GD, et al. Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. JAMA 2008;300:197.
3. Lovell M, Harris K, Forbes T, et al. Peripheral arterial disease: lack of awareness in Canada. Can J Cardiol 2009;25:39-45.
4. Moussa ID, Jaff MR, Mehran R, et al. Prevalence and prediction of previously unrecognized peripheral arterial disease in patients with coronary artery disease: the Peripheral Arterial Disease in Interventional Patients Study. Catheter Cardiovasc Interv 2009;73:719-24.
5. Eikelboom JW, Connolly SJ, Bosch J, et al. Rivaroxaban with or without aspirin in stable cardiovascular disease. N Engl J Med 2017;377:1319-30.
6. Hong Y, Sebastianski M, Makowsky M, Tsuyuki R, McMurtry MS. Administrative data are not sensitive for the detection of peripheral artery disease in the community. Vasc Med 2016;21:331-6.
7. Gouda P, Ramasundarahettige C, Anand S, et al. Clinical factors associated with peripheral artery disease in patients with documented coronary artery disease: a post hoc analysis of the COMPASS trial. Atherosclerosis 2021;331:38-44.
8. Uno H, Cai T, Tian L, Wei L-J. Evaluating prediction rules for t-year survivors with censored regression models. J Am Stat Assoc 2007;102:527-37.

Supplementary Material

To access the supplementary material accompanying this article, visit *CJC Open* at https://www.cjcopen.ca/ and at https://doi.org/10.1016/j.cjco.2023.06.003.