

Validation of Medication Proxies for the Identification of Hospitalizations for Major Adverse Cerebro-Cardiovascular Events

Dennis Steenhuis , Xuechun Li , Talitha L Feenstra, Eelko Hak 

Groningen Research Institute of Pharmacy, Unit of Pharmacotherapy, Epidemiology & Economics, University of Groningen, Groningen, the Netherlands

Correspondence: Dennis Steenhuis, PhD Research Fellow, Groningen Research Institute of Pharmacy, Unit of Pharmacotherapy, Epidemiology and Economics, University of Groningen, A. Deusinglaan 1, Groningen, 9713 AV, the Netherlands, Tel +31 6 8200 4471, Email d.steenhuis@rug.nl

Purpose: To date, complete and long-term registrations of diseases and events are not available in every situation. As a useful proxy, medication usage data is very promising. For instance, real-world dispensing data from pharmacies are attractive because of the high validity of drug dispensing data, large sample sizes, and long-term registration. However, before application as a proxy, validity must be assessed. Therefore, in this study, we aim to assess the validity of various medicines used as a proxy for major adverse cerebrovascular events (MACCE), that is, to identify an incident or previous hospitalization for a MACCE.

Patients and Methods: Using the claims database of a large Dutch healthcare insurer, we estimated the concordance between hospitalization claims for MACCE and specific claims for dispensings to treat MACCE in a cohort of patients on primary preventive antihypertensive and/or antihyperlipidemic therapy between 2013 and 2020.

Results: In a cohort of more than 110,000 patients, a dispensing of either vitamin K antagonists, platelet aggregation inhibitors, or nitrates was predictive of an incident hospitalization for a MACCE between 2013 and 2020, with a sensitivity of 71.5% (95% CI: 70.4–72.5%) and specificity of 93.2% (95% CI 91.1–93.4%), and any history of hospitalization for a MACCE (prevalence) with a sensitivity of 86.9% (95% CI: 86.5–87.3%) and specificity of 81.9 (95% CI: 81.6–82.1%), while positive predicted value remains low. Sensitivity analyses across age, sex, and patients with asthma/COPD or diabetes showed a similarly good performance.

Conclusion: Claims for the dispensings of vitamin K antagonists, platelet aggregation inhibitors, and/or nitrates can be validly used as a proxy for new and previous hospitalization for MACCE in patients on primary preventive therapy. Further study is required to assess the validity of such dispensing claims for non-hospitalized cerebro-cardiovascular events and whether the results are generalizable in non-Dutch cohorts.

Keywords: health insurance claims, cardiovascular events, sensitivity and specificity, pharmaceutical cost groups

Introduction

One of the aims of pharmaco-epidemiology is to estimate the real-world beneficial or side effects of medications.¹ Although randomized controlled trials (RCTs) are seen as the golden standard for causal claims, the use of a RCT is not suitable or possible for every pharmaco-epidemiological research question.² Therefore, the use of big data from secondary healthcare databases, also described as routine care data, is common.³ The use of secondary databases has some advantages over RCTs, such as the presence of real-world users which may differ from clinical trial study populations, larger sample sizes, lower costs, longer follow-up and thus increased generalizability.² Since secondary databases are not necessarily designed for research purposes,³ clinically relevant factors for answering specific research questions may be absent.² The use of secondary database is therefore always a trade-off between the number of available observations and amount and validity of available information.⁴

To enhance the validity of a database for research purposes, data can be enriched. Enrichment is possible by combining the main database with other data sources or by using available data as well as possible. A frequently applied

example of database enrichment is the use of medication proxies to indicate a patients' medical conditions. Examples of such proxies include flagging insulin and/or oral hypoglycemic agent users as diabetes patients,⁵ or use of drugs for chronic airway diseases to identify patients with asthma or COPD.⁶ Similar proxies have been used for several other diseases such as cardiovascular diseases, HIV, and Parkinson's disease.^{7,8} These proxies were used to estimate the prevalence of certain diseases^{9,10} or in risk-adjustment systems.¹¹ Most of which focus on chronic conditions, in contrast to acute events. For example, Füssenich et al¹² showed that stroke was more difficult to predict with pharmaceutical prescriptions than several chronic conditions.

Since cerebro-cardiovascular diseases were both globally and in Western Europe, the leading cause of death in 2019, with an estimated 33% of all deaths both globally and in Western Europe,^{13,14} research towards preventive medication for these diseases as well as these diseases as an outcome themselves is needed. Unfortunately, not every secondary database contains the desired outcome; hence, enrichment of these databases using proxies is an important method for investigating the relationship between available information in the database and an unavailable outcome. Therefore, a validated medication proxy for both history of, and incident cerebro-cardiovascular events is important to support studies and findings based on medication-use databases.

In this study, we will replicate the validation study by Pouwels et al,¹⁵ who presented the accuracy of several medication proxies for the identification of incident and prevalent major cerebro-cardiovascular events. The study by Pouwels et al¹⁵ was based on data from 2008 and 2009 of approximately 17,000 patients with type 2 diabetes. We focused on the same medication classes as Pouwels et al¹⁵ to recreate the same type of results. Next to that, these medication classes are shown to be clinically relevant as they are found among have the highest prevalence in patients with cerebro-cardiovascular diseases by Ma et al.¹⁶

Access to the large healthcare insurance claims database provided us the opportunity to repeat the analysis with more recent data from 2013 to 2021 and for a larger population of primary preventive patients starting any anti-hypertensive or anti-lipidemic drug. The aim is to examine the accuracy of the proposed medication proxies. Subgroup analyses were planned according to age, sex, and relevant comorbidities such as diabetes.

Methods

Study Design and Data Source

Using the claims database of a large Dutch healthcare insurer, we conducted a retrospective population-based inception cohort study. This database contains claims from at least two million people over more than five years and covers a representative sample of about 12% of the Dutch population.¹⁷ For the time someone is insured, the database, except for over-the-counter medication, is complete regarding pharmaceutical dispensings and hospital admissions. To enhance transparency for patients, healthcare providers, and health insurers, the Dutch diagnose-behandel-combinatie (diagnosis-treatment-combination; DBC) system is developed for managing declarations. Unfortunately, there is no direct mapping from this system to international coding systems, such as the International Classification of Diseases tenth revision (ICD-10), so a diagnosis is not available in this type of claims data. Nevertheless, we tried to add the closest related ICD-10 code in [Supplementary Table A](#). Due to the large number of participants in combination with de-identified data, informed consent was not required.¹⁸

Study Population

We used the same dataset as in a previous study¹⁹ in which we included adult patients who started any antihypertensive and/or antihyperlipidemic therapy anywhere between 2013 and 2020. The index date was defined as the date of the initial claim for dispensing antihyperlipidemic or antihypertensive drugs. To be included in this cohort study, patients were required to have a minimum of two years of medical history in the claims database prior to the index date and at least two claims for dispensing antihyperlipidemic or antihypertensive drugs within one year after the index date, and patients were required to be 19 years or older on the index date. Patients were followed for a maximum of 10 years, with the end of the follow-up period, patient death, or patient stopping insurance, whatever came first, indicating the end of follow-up. More information about the inclusion and exclusion criteria and the definition of antihyperlipidemic and antihypertensive therapy can be found in Steenhuis et al.¹⁹

Cerebro-Cardiovascular Drugs

In this study, we used claims for dispensing the following cerebro-cardiovascular drugs as a proxy for a hospitalization for cerebro-cardiovascular events: Vitamin K antagonists (with ATC code starting with B01AA), platelet aggregation inhibitors (B01AC), and nitrates (C01DA). Each of these three classes were tested separately, as well as the combination of Vitamin K antagonists and platelet aggregation inhibitors; and the combination of all the three abovementioned medication classes. In the case of a combination of drugs, the first claim for dispensing one of the drug classes was used as an identifier for a cerebro-cardiovascular event.

Identification of Hospitalization for Incident MACCE

We aimed to evaluate whether we can validly identify patients with an incident major acute cerebro-cardiovascular event using different claims for dispensing drugs for the treatment of cerebro-cardiovascular diseases. For this purpose, the first MACCE hospitalization (based on a claim in [Supplementary Table A](#)) was defined as an incident MACCE. All first MACCE hospitalizations after the index date were defined as incident cases. We were interested in cerebro-cardiovascular drug treatment initiation, defined as the first claim of dispensing a cerebro-cardiovascular drug with no such claim of dispensing that drug in the 730 days before the first dispensing. Patients who already had at least one claim of dispensing the drug of interest or had a MACCE hospitalization in the 730 days before or 90 days after their index date were excluded from this analysis. Drugs dispensed for the first time between 30 days before and 90 days after the MACCE hospitalization event were considered true positives, when dispensed more than 30 days before or more than 90 days after the MACCE hospitalization or when there was no hospitalization as a false positive, and when not dispensed during the entire follow-up period, while there was a MACCE hospitalization as a false negative. Identification window of 30 days before and 90 days after the incident MACCE hospitalization were set equal to the study of Pouwels et al.¹⁵

Identification of History of Hospitalization for MACCE

The outcome was defined as the first MACCE hospitalization (based on a claim in [Supplementary Table A](#)) after 2012. Drug dispensings 30 days before or any time after the MACCE hospitalization event were considered as true positives, when dispensed more than 30 days before the hospitalization or when there was no hospitalization as a false positive, and when not dispensed while there was a hospitalization as a false negative. Patients who left the database within two years of their first hospitalization (by either leaving the health insurer or death) were excluded from the analysis. Patients with a cerebro-cardiovascular drug dispensing or MACCE hospitalization before the defined outcome were excluded from the analysis. Identification window of 30 days before and unlimited time after the MACCE hospitalization were set equal to the study of Pouwels et al.¹⁵

Statistical Analysis

We analyzed the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, area under receiver operating characteristic curve (AUROC)²⁰ and Cohen's Kappa^{20,21} of medications relative to history of or incident MACCE hospitalizations. We chose to report all accuracy measures, as the priority of accuracy measures largely depends on the type of research.²² Confidence intervals for the sensitivity, specificity, PPV, NPV, and accuracy were calculated using the exact method of Clopper and Pearson,²³ with a significance level of 5%. All calculations were performed using SAS Enterprise Guide 9.4.

Sensitivity Analyses

The above-mentioned analyses were repeated for the following subgroups: patients who started antihyperlipidemic therapy only, patients who received antihypertensive therapy only, patients who started both antihypertensive and antihyperlipidemic therapy, patients with diabetes, patients with asthma/COPD, men, women, and three different age groups. Patients who received at least 180 DDD of blood glucose-lowering drugs (ATC code starting with A10) within a year or had a hospital claim related to diabetes at any time during the study period were classified as having diabetes ([Supplementary Table B1](#)). Similarly, patients were classified as having asthma/COPD when they received at least 180

DDD of glucocorticoids (R03BA), adrenergics in combination with corticosteroids (R03AK), adrenergics in combination with anticholinergics (R03AL) within a year, or had a hospital claim related to asthma and/or COPD ([Supplementary Table B1](#)). Selection of subgroups was based on variables that can add to cardiovascular risk.¹⁴

Furthermore, to identify incident hospitalizations for MACCE, we performed a sensitivity analysis by varying the identification window. Varying the number of days before the hospitalization with 45, 15 and 0 and the number of days after the event with 60 and 120 days.

Results

Identification of Incident Hospitalization for MACCE

A total of 113,896 patients receiving primary preventive antihypertensive and/or antihyperlipidemic therapy were selected for analysis. More than half (52.2%) of the study population was male, with a mean age of 58.1 years (SD: 13.0). A total of 5.6% were diagnosed with diabetes and 8.1% were diagnosed with asthma/COPD. Of the 113,896 patients, 31,383 (27.6%) started antihyperlipidemic therapy, 77,947 (68.4%) started antihypertensive therapy, and 4,566 (4.0%) started both hyperlipidemic and antihypertensive therapy. In total, 10,965 patients (9.6%) were hospitalized for MACCE after the index date ([Table 1](#)).

Three different medication proxies (Vitamin K antagonists, platelet aggregation inhibitors, and nitrates) for the identification of a hospitalization for an incident MACCE had a sensitivity ranging from 2.2% to 66.2% and a PPV ranging from 13.7% to 50.2%. Specificity and NPV for the three different proxies were more similar (94.6–98.6% resp. 90.9–97.2%). Combining the three different proxies increased the sensitivity to 71.5% (95% confidence interval [95% CI]: 70.4–72.5%), whereas the PPV was 44.9% (95% CI: 44.0–45.8, [Table 2](#)). Platelet aggregation inhibitors were the most important proxy.

Identification of History of MACCE Hospitalization

For the second analysis, we selected a total of 136,703 patients who started antihypertensive and/or antihyperlipidemic therapy. Demographics of this patient group were similar to those of the previous analysis. The largest difference was in the percentage of patients who started antihypertensive therapy or the combination of antihypertensive and antihyperlipidemic therapy. In total, 25.4% of the patients had a recorded history of a hospitalization for a cerebro-cardiovascular event ([Table 1](#)).

Table 1 Baseline Characteristics of Both Patients' Groups That Were Used for Analysis

	Cohort for Identification of the Hospitalization for an Incident MACCE	Cohort for Identification of the History of Any Hospitalization for a MACCE
N	113,896	136,703
Age, mean, years (SD)	58.1 (13)	58.6 (12.9)
Men, n (%)	59,428 (52.2%)	74,924 (54.8%)
Diabetic, n (%)	6,425 (5.6%)	7,071 (5.2%)
Asthma/COPD, n (%)	9,221 (8.1%)	11,069 (8.1%)
Antihyperlipidemic, n (%)	31,383 (27.6%)	38,939 (28.5%)
Antihypertensive, n (%)	77,947 (68.4%)	82,085 (60.0%)
Both, n (%)	4,566 (4.0%)	15,679 (11.5%)
MACCE after indexdate, n (%)	10,965 (9.6%)	
Any MACCE, n (%)		34,756 (25.4%)

Abbreviations: SD, Standard Deviation; COPD, Chronic obstructive pulmonary disease; MACCE, Major Adverse Cerebro-Cardiovascular event.

Table 2 Identifying Incident Hospitalizations for an MACCE

Drug Prescription	Vitamin K Antagonists	Platelet Aggregation Inhibitors	Nitrates	Vitamin K Antagonists / Platelet Aggregation Inhibitors	Vitamin K Antagonists / Platelet Aggregation Inhibitors / Nitrates
True Positives	219	5,128	1,901	5,067	5,058
False Positives	1,381	5,097	2,145	5,713	6,211
True Negatives	99,147	89,295	100,272	86,337	85,581
False Negatives	9,868	2,616	7,756	2,177	2,019
Sensitivity (95% CI)	2.2 (1.9–2.5)	66.2 (65.2–67.3)	19.7 (18.9–20.5)	69.9 (68.9–71)	71.5 (70.4–72.5)
Specificity (95% CI)	98.6 (98.6–98.7)	94.6 (94.5–94.7)	97.9 (97.8–98)	93.8 (93.6–93.9)	93.2 (93.1–93.4)
Accuracy (95% CI)	89.8 (89.7–90)	92.4 (92.3–92.6)	91.2 (91–91.3)	92.1 (91.9–92.2)	91.7 (91.5–91.8)
Kappa (95% CI)	0.01 (0–0.03)	0.53 (0.52–0.54)	0.24 (0.22–0.25)	0.52 (0.51–0.53)	0.51 (0.5–0.52)
AUROC (95% CI)	0.5 (0.5–0.51)	0.8 (0.8–0.81)	0.59 (0.58–0.59)	0.82 (0.81–0.82)	0.82 (0.82–0.83)
PPV (95% CI)	13.7 (12–15.5)	50.2 (49.2–51.1)	47 (45.4–48.5)	47 (46.1–48)	44.9 (44–45.8)
NPV (95% CI)	90.9 (90.8–91.1)	97.2 (97–97.3)	92.8 (92.7–93)	97.5 (97.4–97.6)	97.7 (97.6–97.8)

Abbreviations: AUROC, Area Under Receiver Operating Characteristic curve; P/(N)PV, Positive/(Negative) Predictive Value.

The three different medication proxies showed a sensitivity ranging from 5.2% to 83.9% and a PPV ranging from 29.8% to 72.9%. The specificity and NPV showed ranges that were closer together but wider than those used for the identification of hospitalizations for incident MACCE. Combining the three proxies increased the sensitivity to 86.9% (95% CI: 86.5–87.3%) (Table 3).

Sensitivity Analyses

All results of the sensitivity analyses pointed in the same direction as the main analyses. We noted a difference in the sensitivity of the proxy for hospitalization for an incident MACCE in patients who started antihyperlipidemic versus

Table 3 Identifying History of a Hospitalization for a MACCE

Drug Prescription	Vitamin K Antagonists	Platelet Aggregation Inhibitors	Nitrates	Vitamin K Antagonists / Platelet Aggregation Inhibitors	Vitamin K Antagonists / Platelet Aggregation Inhibitors / Nitrates
True Positives	1,721	25,153	10,153	25,099	25,025
False Positives	3,795	13,908	3,607	16,624	17,529
True Negatives	97,770	87,906	99,019	84,721	83,881
False Negatives	31,916	4,886	23,294	4,057	3,818
Sensitivity (95% CI)	5.1 (4.9–5.4)	83.7 (83.3–84.2)	30.4 (29.9–30.9)	86.1 (85.7–86.5)	86.8 (86.4–87.2)
Specificity (95% CI)	96.3 (96.1–96.4)	86.3 (86.1–86.6)	96.5 (96.4–96.6)	83.6 (83.4–83.8)	82.7 (82.5–82.9)
Accuracy (95% CI)	73.6 (73.4–73.8)	85.7 (85.6–85.9)	80.2 (80–80.4)	84.2 (84–84.4)	83.6 (83.4–83.8)
Kappa (95% CI)	0.02 (0.01–0.03)	0.63 (0.63–0.64)	0.33 (0.33–0.34)	0.6 (0.6–0.61)	0.59 (0.59–0.6)
AUROC (95% CI)	0.51 (0.51–0.51)	0.85 (0.85–0.85)	0.63 (0.63–0.64)	0.85 (0.85–0.85)	0.85 (0.85–0.85)
PPV (95% CI)	31.2 (30–32.4)	64.4 (63.9–64.9)	73.8 (73–74.5)	60.2 (59.7–60.6)	58.8 (58.3–59.3)
NPV (95% CI)	75.4 (75.2–75.6)	94.7 (94.6–94.9)	81 (80.7–81.2)	95.4 (95.3–95.6)	95.6 (95.5–95.8)

Abbreviations: AUROC, Area Under Receiver Operating Characteristic curve; P/(N)PV, Positive/(Negative) Predictive Value.

Table 4 Sensitivity Analysis for Identifying Incident Hospitalizations for a MACCE

	TP	FP	TN	FN	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Kappa (95% CI)	AUROC (95% CI)
Total	5,058	6,211	85,581	2,019	71.5 (70.4–72.5)	93.2 (93.1–93.4)	91.7 (91.5–91.8)	0.51 (0.5–0.52)	82.4 (81.9–82.8)
Starting medication									
Antihyperlipidemics	1,292	1,450	22,451	387	77 (74.9–78.9)	93.9 (93.6–94.2)	92.8 (92.5–93.1)	0.55 (0.53–0.57)	85.4 (84.5–86.3)
Antihypertensives	3,576	4,531	60,631	1,567	69.5 (68.3–70.8)	93 (92.8–93.2)	91.3 (91.1–91.5)	0.49 (0.48–0.51)	81.3 (80.7–81.8)
Comorbidities									
Diabetes	1,004	1,055	12,134	255	79.7 (77.4–81.9)	92 (91.5–92.5)	90.9 (90.5–91.4)	0.56 (0.53–0.58)	85.9 (84.9–86.7)
COPD	776	1,167	9,383	325	70.5 (67.7–73.2)	88.9 (88.3–89.5)	87.2 (86.6–87.8)	0.44 (0.42–0.47)	79.7 (78.6–80.7)
Sex									
Men	3,117	3,296	42,939	1,088	74.1 (72.8–75.4)	92.9 (92.6–93.1)	91.3 (91.1–91.6)	0.54 (0.53–0.55)	83.5 (82.9–84)
Women	1,941	2,915	42,642	931	67.6 (65.8–69.3)	93.6 (93.4–93.8)	92.1 (91.8–92.3)	0.46 (0.45–0.48)	80.6 (79.8–81.3)
Age									
< 50	742	1,115	24,037	419	63.9 (61.1–66.7)	95.6 (95.3–95.8)	94.2 (93.9–94.5)	0.46 (0.44–0.49)	79.7 (78.4–81)
50–69	1,269	1,537	13,591	483	72.4 (70.3–74.5)	89.8 (89.3–90.3)	88 (87.5–88.5)	0.49 (0.47–0.51)	81.1 (80.3–81.9)
70+	3,047	3,559	47,953	1,117	73.2 (71.8–74.5)	93.1 (92.9–93.3)	91.6 (91.4–91.8)	0.52 (0.51–0.54)	83.1 (82.6–83.7)
Identification window									
(–45,90)	5,151	6,118	85,581	2,019	71.8 (70.8–72.9)	93.3 (93.3–93.5)	91.8 (91.6–91.9)	0.52 (0.51–0.53)	82.6 (82.1–83)
(–15, 90)	4,881	6,388	85,581	2,019	70.7 (69.6–71.8)	93.1 (93.1–93.2)	91.5 (91.3–91.7)	0.49 (0.48–0.5)	81.9 (81.4–82.3)
(0, 90)	4,181	7,088	85,581	2,019	67.4 (66.3–68.6)	92.4 (92.4–92.5)	90.8 (90.6–91)	0.43 (0.42–0.44)	79.9 (79.4–80.4)
(–30,60)	5,113	6,684	85,581	2,094	70.9 (69.9–72)	92.8 (92.8–92.9)	91.2 (91–91.4)	0.49 (0.48–0.5)	81.9 (81.4–82.3)
(–30,120)	4,943	5,886	85,581	1,962	71.6 (70.5–72.6)	93.6 (93.6–93.7)	92 (91.9–92.2)	0.52 (0.51–0.53)	82.6 (82.1–83)
(–45,120)	5,030	5,799	85,581	1,962	71.9 (70.9–73)	93.7 (93.7–93.8)	92.1 (91.9–92.3)	0.52 (0.51–0.53)	82.8 (82.3–83.2)
(0,60)	4,204	7,593	85,581	2,094	66.8 (65.6–67.9)	91.9 (91.9–92)	90.3 (90.1–90.4)	0.42 (0.41–0.43)	79.3 (78.8–79.8)

Abbreviations: TP, True Positives; FP, False Positives; TN, True Negatives; FN, False; AUROC, Area Under Receiver Operating Characteristic Curve.

antihypertensive therapy. A higher sensitivity was observed in the group of patients with diabetes (Table 4). Next to that, we noted a lower sensitivity, specificity, PPV, NPV, and accuracy in all sensitivity analyses in which the identification window before the event was set to zero days. The results of other proxies and those of the proxy for any history of hospitalization for a cardiovascular event can be found in [Supplementary Tables C and D](#)).

Discussion

In the group of patients starting primary preventive antihypertensive and/or antihyperlipidemic drug therapy, more than 70% of the hospitalizations for incident MACCE and 86% of a history of such hospitalizations could be identified using the proxy of a single claim for a dispensing of either platelet aggregation inhibitors, nitrates, or vitamin K antagonists.

Our three medication classes (ATC: B01AA, B01AC, and C01D) contained the drugs found as having the highest prevalence (excluding the drugs we defined as antihypertensive and antihyperlipidemic therapy) in patients with cerebrocardiovascular diseases by Ma et al.¹⁶ Based on their research, the addition of furosemide (C03CA01) or the complete class of drugs plain sulfonamides (C03CA) could be an addition to our best medication proxy for identifying cardiovascular events. Less than 80 patients or 4% of the false negatives could be identified using plain sulfonamides. Although

looking at a Scottish and slightly different study population (patients not necessarily on antihypertensive and/or antihyperlipidemic therapy), the results of Payne et al²⁴ pointed in the same direction and found a difference in sensitivity between nitrates and platelet aggregation inhibitors of the same order. Finally, compared to Pouwels et al¹⁵ we found approximately the same sensitivity and specificity. Both the PPV and NPV were significantly higher than those reported by Pouwels et al¹⁵ but this could be explained by different study populations. Since they used a diabetes-only population, the incidence and prevalence of MACCE will differ, which affects both the PPV and NPV.^{25,26} This even holds when looking at a diabetes-only population (Table 4 and [Supplementary Table C](#)), since we examined a group of patients with diabetes using primary preventive cerebro-cardiovascular therapy (antihyperlipidemic and/or antihypertensive medication), while Pouwels et al¹⁵ studied an unselected group of patients with type II diabetes.

Using our best medication proxy, we correctly identified 71.5% of the incident hospitalizations for MACCE. The inability to identify the other 28.5% could have several reasons: Firstly, patients died, which may or may not be a result of the MACCE, or left the database before they were able to recover and pick up a prescription. In our dataset, we could not determine the cause of death for those who died before hospitalization, but only less than 1% of the patients who were hospitalized for a cerebro-cardiovascular event (our cases) died within 90 days after the hospitalization, while another less than 1% of the patients with a hospitalization for a MACCE left the database within 90 days. Hence, this does not explain the 28.5% unidentifiable cases. Secondly, patients were prescribed other types of cerebro-cardiovascular drug therapy. After an extensive search for any other types of medications prescribed in case of a hospitalization for a cerebro-cardiovascular event, only proton pump inhibitors (A02BC) appeared more frequently; however, these drugs are used for gastric protection when aspirin is prescribed²⁷ and hence do not qualify as a proxy. Furthermore, direct acting oral anticoagulants (B01AF) could be an addition to the proxy as the next cerebro-cardiovascular event related medication in the list, but they were only found in less than 15% of the false negatives. Next to that, the classes of medication our study population was based on could be a possible identifier for hospitalization for cardiovascular events. Therefore, statins could be an identifier for hospitalization for a cardiovascular event in cases where the patient is on antihypertensive therapy ([Supplementary Table E](#)), or vice versa. Using antihyperlipidemic or antihypertensive therapy as a proxy in our dataset could identify approximately 2.8% of incident MACCE hospitalizations. Thirdly, patients were not prescribed relevant medications for the treatment of MACCE. This could be the case when patients experience a less severe cerebro-cardiovascular event, such as a TIA.

Up to 50.2% of the patients (depending on the proxy) who were classified as a MACCE hospitalization were actually hospitalized. Although the selected medication proxies are mainly prescribed in the case of a cardiovascular event, not all cardiovascular events are severe enough or noticed in a timely manner such that a patient goes to a hospital within the time window we set. Hence, these claims for a dispensing could indicate a MACCE which did not lead to a hospitalization.

In the sensitivity analysis, a higher sensitivity was observed for patients with diabetes, starting with antihyperlipidemic medication and men. This could be a result of subgroups already having a higher risk of suffering from a MACCE.^{28,29} A lower sensitivity and AUROC scores can be found in the subgroups with an age less than 50, women and those with COPD, which could be a result of a lower risk of suffering from a MACCE.^{28,29} Next to that, lower sensitivity, specificity, PPV, NPV, and accuracy were observed when the identification window before the hospitalization was set to zero. This indicates that patients were prescribed medication for the treatment before the actual event and could be a result of medication being prescribed preoperatively.

To determine the most useful proxy, sensitivity and specificity values for each proxy were added. Based on a cut-off value of 1.5 for sensitivity + specificity,²⁶ we conclude that the proxies based on platelet aggregation inhibitors only; platelet aggregation inhibitors and/or vitamin K antagonists; or platelet aggregation inhibitors and/or vitamin K antagonists and/or nitrates were all useful proxies for identifying incident MACCE hospitalizations as well as a history of such hospitalization. Kappa values for the abovementioned proxies for both hospitalization for incident MACCE as well as any history of MACCE hospitalization ranged from 0.51 to 0.6, indicating a moderate agreement between the proxy and events.³⁰ This was also true for all investigated subgroups. AUROC scores for these three proxies for both outcomes were above 0.8, indicating a very good diagnostic accuracy.³¹ Furthermore, all our proxies had a specificity over 93%, which indicates that the association between our proxies and the outcome is less likely due to other variables, hence a larger probability of

a causal relationship between proxy and outcome (or the other way around) can be assumed.³² The PPV of all proxies for the identification of an incident hospitalization for a MACCE is low. Low PPV follows from a low number of true positives in relation to false positives. Higher PPV can be seen if medication dispensings can be used as a identifier for a GP diagnoses as well.¹⁵ The results in this study are based on Dutch claims database. Therefore, one should be careful in generalizing these results towards other countries, since different prescribing guidelines could cause misclassification and less access to pharmacies can underidentify severe cases of MACCE. Nevertheless, the results of this study can be used in database in which only pharmaceutical data is available to enhance the data.

The medication classes used in this research partly overlapped with some pharmaceutical cost groups used in the Dutch risk-adjustment system.³³ However, it should be kept in mind that the latter was intended to predict healthcare costs of the chronically ill, while we focused on the identification of an event.

Using more advanced statistical techniques or adding variables, such as age or sex, could improve the prediction model used. However, this comes with the risk of creating a model that is much more difficult to understand or the need to gather more data than is available. The strength of the designed model is that it can be used in every database where medication use information is available.

Conclusion

The combined medication proxy of any claim for a dispensing of either vitamin K antagonists, platelet aggregation inhibitors, or nitrates could accurately identify 71.5% of the hospitalizations for an incident MACCE in patients on primary preventive antihypertensive and/or antihyperlipidemic therapy, while it could correctly identify 86.9% of the patients on primary preventive antihypertensive and/or antihyperlipidemic therapy who were hospitalized for a cerebro-cardiovascular event before the claim for a dispensing of their medication. Nevertheless, PPV of all proxies was low for identifying a hospitalization for an incident MACCE. Further research is required to determine whether these (or other) proxies can identify cardiovascular events without a hospitalization, or milder cerebro-cardiovascular events such as a TIA as well.

Data Sharing Statement

The data used for this study were provided by Menzis Zorgverzekeraar N.V. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the first author with the permission of Menzis Zorgverzekeraar N.V.

Ethics Approval and Informed Consent

The Menzis Zorgverzekeraar N.V. claims database contains data that is collected in accordance with Dutch Code of Conduct for Processing Personal Data by Health Insurers, to which Dutch and European (GDPR) privacy legislation applies. According to the Dutch Central Committee on Research Involving Human Subjects (CCMO), approval from the medical ethics committee was not necessary nor required for this study being a retrospective cohort study.³⁴ Informed consent was waived by using deidentified administrative data with a large number of participants in a retrospective design.

Funding

Li is funded by the China Scholarship Council (file no: 202106070028). The grant agency does not impose restrictions on conduct of analyses or dissemination of findings.

Disclosure

Steenhuis reports personal fees from Menzis Zorgverzekeraar N.V., outside the submitted work. Li is funded by the China Scholarship Council (file no: 202106070028). Feenstra and Hak do not declare any conflicts of interest in this work.

References

1. Montastruc J-L, Benevent J, Montastruc F, et al. What is pharmacoepidemiology? Definition, methods, interest and clinical applications. *Therapies*. 2019;74(2):169–174. doi:10.1016/j.therap.2018.08.001
2. Harpe SE. Using secondary data sources for pharmacoepidemiology and outcomes research. *Pharmacotherapy*. 2009;29(2):138–153. doi:10.1592/phco.29.2.138
3. Gandhi S, Salmon JW, Kong SX, Zhao SZ. Administrative databases and outcomes assessment: an overview of issues and potential utility. *J Managed Care Pharm*. 1999;5(3):215–222. doi:10.18553/jmcp.1999.5.3.215
4. Wray NP, Ashton CM, Kuykendall DH, et al. Using administrative databases to evaluate the quality of medical care: a conceptual framework. *Soc Sci Med*. 1995;40(12):1707–1715. doi:10.1016/0277-9536(94)00275-x
5. Lipscombe LL, Hwee J, Webster L, et al. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res*. 2018;18(1):316. doi:10.1186/s12913-018-3148-0
6. Pont LG, van der Werf GT, Denig P, et al. Identifying general practice patients diagnosed with asthma and their exacerbation episodes from prescribing data. *Eur J Clin Pharmacol*. 2002;57(11):819–825. doi:10.1007/s00228-001-0395-4
7. Huber CA, Szucs TD, Rapold R, et al. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. *BMC Public Health*. 2013;13:1030. doi:10.1186/1471-2458-13-1030
8. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol*. 1992;45(2):197–203. doi:10.1016/0895-4356(92)90016-G
9. Slobbe LCJ, Füssenich K, Wong A, et al. Estimating disease prevalence from drug utilization data using the Random Forest algorithm. *Eur J Public Health*. 2019;29(4):615–621. doi:10.1093/eurpub/cky270
10. Chini F, Pezzotti P, Orzella L, et al. Can we use the pharmacy data to estimate the prevalence of chronic conditions? A comparison of multiple data sources. *BMC Public Health*. 2011;11:688. doi:10.1186/1471-2458-11-688
11. Lamers LM, van Vliet RCJA. The pharmacy-based cost group model: validating and adjusting the classification of medications for chronic conditions to the Dutch situation. *Health Policy*. 2004;68(1):113–121. doi:10.1016/j.healthpol.2003.09.001
12. Füssenich K, Boshuizen HC, Nielen MMJ, et al. Mapping chronic disease prevalence based on medication use and socio-demographic variables: an application of LASSO on administrative data sources in healthcare in the Netherlands. *BMC Public Health*. 2021;21(1):1039. doi:10.1186/s12889-021-10754-4
13. Roser M. *Causes of Death Globally: What Do People Die From?* Our World in Data; 2021.
14. Mubarik S, Luo L, Naeem S, et al. Epidemiology and demographic patterns of cardiovascular diseases and neoplasms deaths in Western Europe: a 1990–2019 analysis. *Public Health*. 2024;231:187–197. doi:10.1016/j.puhe.2024.04.003
15. Pouwels KB, Voorham J, Hak E, et al. Identification of major cardiovascular events in patients with diabetes using primary care data. *BMC Health Serv Res*. 2016;16:110. doi:10.1186/s12913-016-1361-2
16. Ma -T-T, Wong ICK, Whittlesea C, et al. Initial cardiovascular treatment patterns during the first 90 days following an incident cardiovascular event. *Br J Clin Pharmacol*. 2021;87(3):1043–1055. doi:10.1111/bcp.14463
17. Nijhof E, Romp M, van Wetten N. Inzicht in het overstapeseizoen. In: *Zorgthermometer Verzekerden in Beeld*. Vektis; 2022.
18. Insurance terms and conditions Menzis Basis 2022 [in Menzis 2021].
19. Steenhuis D, Li X, Feenstra T, et al. The association between deductibles and cardiovascular medication adherence: a retrospective inception cohort study. *Drugs - Real World Outcomes*. 2023;11:99–108. doi:10.1007/s40801-023-00397-9
20. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep*. 2024;14(1):6086. doi:10.1038/s41598-024-56706-x
21. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46. doi:10.1177/001316446002000104
22. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65(3):343–349.e2. doi:10.1016/j.jclinepi.2011.09.002
23. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26(4):404–413. doi:10.1093/biomet/26.4.404
24. Payne RA, Abel GA, Simpson CR. A retrospective cohort study assessing patient characteristics and the incidence of cardiovascular disease using linked routine primary and secondary care data. *BMJ Open*. 2012;2(2):e000723. doi:10.1136/bmjopen-2011-000723
25. Parikh R, Mathai A, Parikh S, et al. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45–50. doi:10.4103/0301-4738.37595
26. Power M, Fell G, Wright M. Principles for high-quality, high-value testing. *Evid Based Med*. 2013;18(1):5–10. doi:10.1136/eb-2012-100645
27. Agewall S, Cattaneo M, Collet JP, et al. Expert position paper on the use of proton pump inhibitors in patients with cardiovascular disease and antithrombotic therapy. *Eur Heart J*. 2013;34(23):1708–1713. doi:10.1093/eurheartj/ehd042
28. Steenhuis D, de Vos S, Bos J, et al. Role of traditional cardiovascular risk factors after initiation of statin therapy: a PharmLines Inception Cohort Study. *Cardiovasc Ther*. 2022;2022:6587165. doi:10.1155/2022/6587165
29. Steenhuis D, de Vos S, Bos JHJ, et al. Risk factors for drug-treated major adverse cardio-cerebrovascular events in patients on primary preventive statin therapy: a retrospective cohort study. *Prev Med Rep*. 2023;34:102258. doi:10.1016/j.pmedr.2023.102258
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
31. Šimundić AM. Measures of diagnostic accuracy: basic definitions. *Ejtfc*. 2009;19(4):203–211.
32. Shimonovich M, Pearce A, Thomson H, et al. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol*. 2021;36(9):873–887. doi:10.1007/s10654-020-00703-7
33. Risk Adjustment Policy Rules 2024 [in Dutch: Beleidsregels Risicoverevening 2024], Staatscourant 2023, 28694.
34. Ministerie van Volksgezondheid, Welzijn en Sport. Your research: is it subject to the WMO or not? Investigators | the Central Committee on Research on Involving Human Subjects. 2023. Available from: <https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not>. Accessed March 25, 2025.

Clinical Epidemiology**Dovepress**

Taylor & Francis Group

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>