






Validation of novel identification algorithms for major adverse cardiovascular events in a Japanese claims database

Daisuke Shima PhD¹  | Yoichi Ii PhD² | Shingo Higa MS¹  | Takahide Kohro MD, PhD³ | Satoshi Hoshide MD, PhD⁴  | Ken Kono MD, PhD⁴ | Shigeru Fujimoto MD, PhD⁵ | Satoshi Nijima MD, PhD⁴ | Naoko Tomitani BSc⁴  | Kazuomi Kario MD, PhD⁴ 

¹Pfizer Japan Inc., Tokyo, Japan

²Pfizer R&D Japan G.K., Tokyo, Japan

³Department of Clinical Informatics, Jichi Medical University School of Medicine, Tochigi, Japan

⁴Division of Cardiovascular Medicine, Department of Medicine, Jichi Medical University School of Medicine, Tochigi, Japan

⁵Division of Neurology, Department of Medicine, Jichi Medical University School of Medicine, Tochigi, Japan

Correspondence

Daisuke Shima, Pfizer Japan Inc., 3-22-7, Yoyogi, Shibuya-ku, Tokyo 151-8589, Japan.

Email: Daisuke.Shima@pfizer.com

Kazuomi Kario, Division of Cardiovascular Medicine, Department of Medicine, Jichi Medical University School of Medicine, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan.

Email: kkario@jichi.ac.jp

Funding information

Pfizer Japan Inc

Abstract

Predicting clinical outcomes can be difficult, particularly for life-threatening events with a low incidence that require numerous clinical cases. Our aim was to develop and validate novel algorithms to identify major adverse cardiovascular events (MACEs) from claims databases. We developed algorithms based on the data available in the claims database International Classification of Diseases, Tenth Revision (ICD-10), drug prescriptions, and medical procedures. We also employed data from the claims database of Jichi Medical University Hospital, Japan, for the period between October 2012 and September 2014. In total, we randomly extracted 100 potential acute myocardial infarction cases and 200 potential stroke cases (ischemic and hemorrhagic stroke were analyzed separately) based on ICD-10 diagnosis. An independent committee reviewed the corresponding clinical data to provide definitive diagnoses for the extracted cases. We then assessed the algorithms' accuracy using positive predictive values (PPVs) and apparent sensitivities. The PPVs of acute myocardial infarction, ischemic stroke, and hemorrhagic stroke were low only by diagnosis (81.6% [95% CI 72.5–88.7]; 31.0% [95% CI 22.8–40.3]; and 45.5% [95% CI 34.1–57.2], respectively); however, the PPVs were elevated after adding the prescription and procedure data (87.0% [95% CI 78.3–93.1]; 44.4% [95% CI 32.7–56.6]; and 46.1% [95% CI 34.5–57.9], respectively). When we added event-specific prescription and procedure data to the algorithms, the PPVs for each event increased to 70%–98%, with apparent sensitivities exceeding 50%. Algorithms that rely on ICD-10 diagnosis in combination with data on specific drugs and medical procedures appear to be valid for identifying MACEs in Japanese claims databases.

1 | INTRODUCTION

Major adverse cardiovascular events (MACEs), such as stroke and myocardial infarction (MI), are the most common cause of death

worldwide (the second leading cause in Japan), and their prevention is critical in healthcare.^{1–3} Hypertension, hyperlipidemia, and diabetes mellitus are known risk factors for MACEs, and a number of pharmaceutical agents have been developed to control the risk

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *The Journal of Clinical Hypertension* published by Wiley Periodicals LLC.

factors. To assess the effect of these pharmaceutical interventions on MACEs, clinical trials with large sample sizes and long-term follow-up have been conducted.⁴⁻⁶ There has been growing interest in recent years in outcome assessments with claims databases, which enable effective data collection in pharmacoepidemiological studies, helping key stakeholders make informed decisions to improve healthcare at the individual and population levels.^{7,8} It has also been proposed that information from claims databases could be employed for pharmaceutical studies, especially when the incidence of MACEs is a primary endpoint.^{9,10}

A claims database is a collection of information obtained by health insurers and public programs for reimbursement that includes data on patient information (eg, age, sex), diagnoses, procedures, and prescriptions with long-term follow-up, enabling the identification, tracking, and analysis of national trends in healthcare use, access, quality, outcomes, and costs.^{11,12} Claims data are created for administrative purposes and therefore have certain challenges for outcome assessments, such as evaluating the data quality, combining data from different databases with dissimilar coding systems, and developing appropriate use.¹³ In particular, algorithms that use diagnostic codes, such as those in the International Classification of Diseases, Tenth Revision (ICD-10),¹⁴ can lead to inaccurate research results due to incorrect or incomplete diagnosis data records.^{12,15} Diagnosis code validation using primary medical data and definitive diagnosis is therefore crucial for proper record keeping and quality healthcare results.

In Japan, claims databases of the national health insurance program have accumulated a wealth of healthcare data, including those related to diagnoses, prescriptions, and medical procedures; however, the databases lack data on disease severity.¹⁶ Most claims databases are large, contain individual patient data, and collect data in the same manner across institutions, allowing for long-term follow-up. The Japanese Ministry of Health, Labour, and Welfare has begun the development of a national claims and health checkup database.¹⁷

According to customary Japanese procedures, a provisional diagnosis is frequently recorded for clinical examinations. This record sometimes remains in the claims database, even after a definitive diagnosis has been recorded in a patient's clinical record, which can result in incomplete and inaccurate data with often misclassified diagnoses. Validation studies are therefore essential for claims database research in Japan.

This study aimed to develop and evaluate algorithms for identifying potential cardiovascular events in a Japanese claims database. We employed the claims database (Jichi database) and electrical medical records of Jichi Medical University Hospital, which is a tertiary emergency care facility for patients who require admission to the intensive care unit in one of the representative hospitals in Japan. This database includes a sufficient number of cardiovascular events and is considered appropriate for use in cardiovascular research.

2 | METHODS

2.1 | Ethical approval

The study was approved by the Ethics Review Committee of Jichi Medical University. Given that the patient information in the claims database was anonymized at the time we extracted the data, written informed consent was not required. Information regarding the opting out of data use by third parties was publicly displayed on the website of the Cardiovascular Medicine Division of Jichi Medical University. The authors declare that all supporting data are available within the article.

2.2 | Data source and study population

The Jichi database, an administrative database at Jichi Medical University Hospital (Tochigi, Japan), collects and stores relevant medical data, including diagnoses with ICD-10 codes, patient background information, drug prescriptions, and medical procedures. We included the records for patients aged at least 20 years who developed incident cardiovascular events requiring medical treatment at Jichi Medical University Hospital between October 1, 2012, and September 30, 2014. There were no exclusion criteria. The cardiovascular events of interest were acute MI, ischemic stroke (excluding transient ischemic attack), and hemorrhagic stroke. To confirm true-positive events, an independent clinical event committee (CEC) utilized the electrical medical records at Jichi Medical University Hospital.

2.3 | Study design

We employed predefined algorithms based on the Jichi database to identify potential incident cardiovascular events. The algorithms consisted of common variables available in other general Japanese claims databases, including diagnosis (ICD-10 codes), diagnosis status (definitive or provisional), prescribed drugs (name, code, dosage, administration, and date prescribed), and medical procedures (name, code, hospitalization, and date).

Potential MI events were identified by ICD-10 code I21 (acute MI) or I22 (subsequent MI); drugs were classified as anticoagulant, heparin, anti-thrombin, aspirin, or thrombolytic agent; and medical procedures were classified as coronary angioplasty, myocardial marker tests (troponin T/I, creatinine kinase [CK], and CK-MB type), or hospitalization. Potential ischemic stroke events were identified by ICD-10 code I63 (cerebral infarction); drugs were classified as antiplatelet drugs, anticoagulants, heparin, anti-thrombin, aspirin, cerebral metabolism activator, or thrombolytic agents; and medical procedures were classified as head imaging (computed tomography [CT] or magnetic resonance imaging [MRI]), carotid artery

ultrasonography, hospitalization, or rehabilitation. Potential hemorrhagic stroke events were identified by ICD-10 code I60 (subarachnoid hemorrhage), I61 (intracerebral hemorrhage), or I62 (other non-traumatic intracranial hemorrhage); no drugs were applicable for the selection of cases; and medical procedures were classified as head imaging (MRI/CT), hospitalization, or rehabilitation. For each cardiovascular event, we included all drugs and medical procedures recorded in the month of diagnosis or in the subsequent month.

In the primary analysis, the algorithms were developed based on definitive ICD-10 diagnoses only, diagnoses plus drugs, diagnoses plus medical procedures, and diagnoses plus drugs plus medical procedures. As exploratory analyses, we also examined the algorithms that employed subgroups of the factors defining the diagnoses, drugs, and medical procedures from the claims database. The

validation study team performed the case extraction, algorithm development, and analysis.

2.4 | Clinical event committee

The CEC, which consisted of 3 physicians (2 cardiologists and 1 neurologist) who were independent of the validation study team, reviewed the electronic medical record data for the extracted potential cases to assess whether the cardiovascular events were true-positive incident cases based on the following definitions. Figure 1 summarizes the process the CEC followed.

An acute MI event was recorded when any of the following conditions were met¹⁸: (a) a CK-MB value ≥ 2 times the upper reference

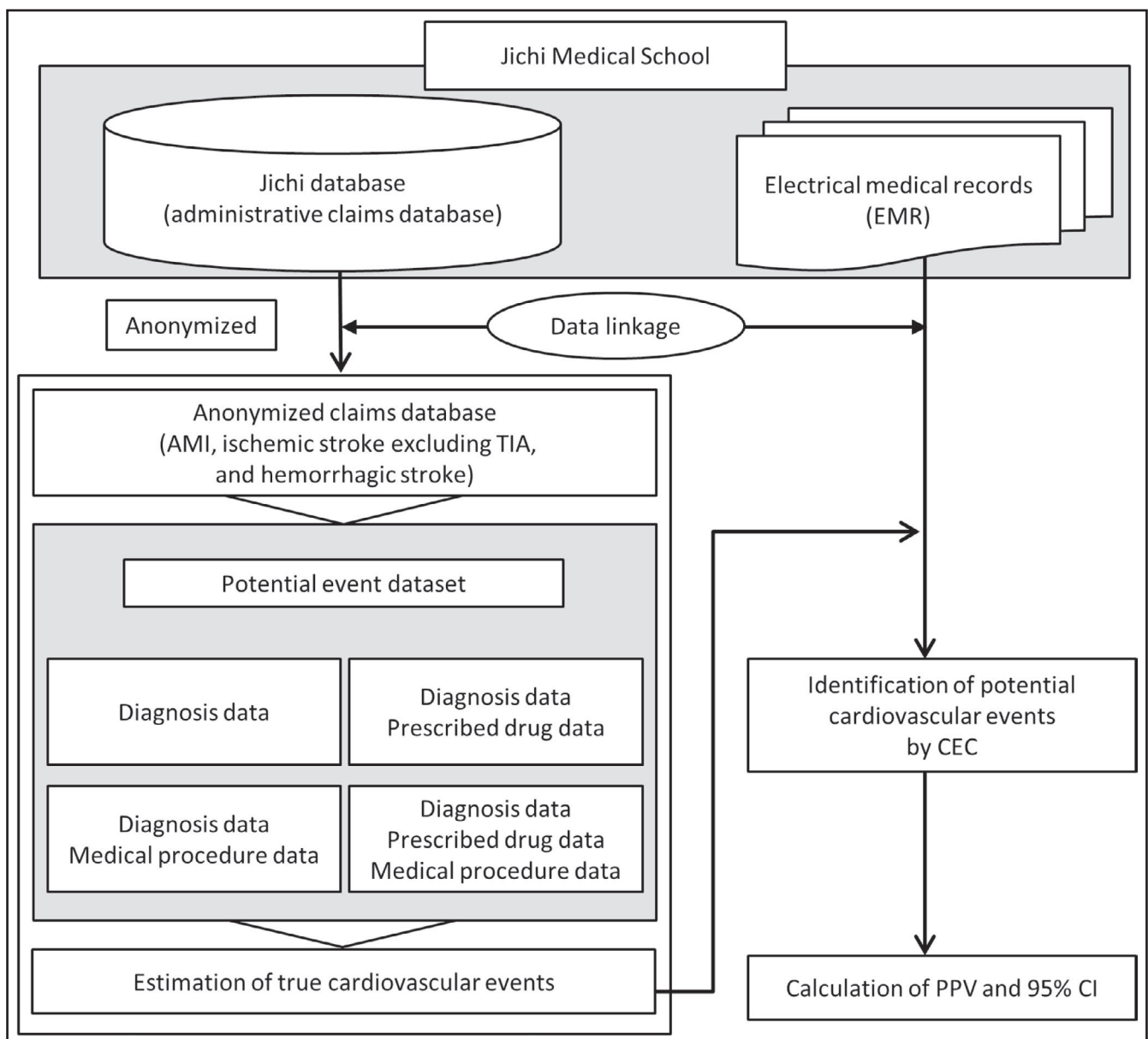


FIGURE 1 Validation process. AMI, acute myocardial infarction; CEC, clinical event committee; PPV, positive predictive value; TIA, transient ischemic attack

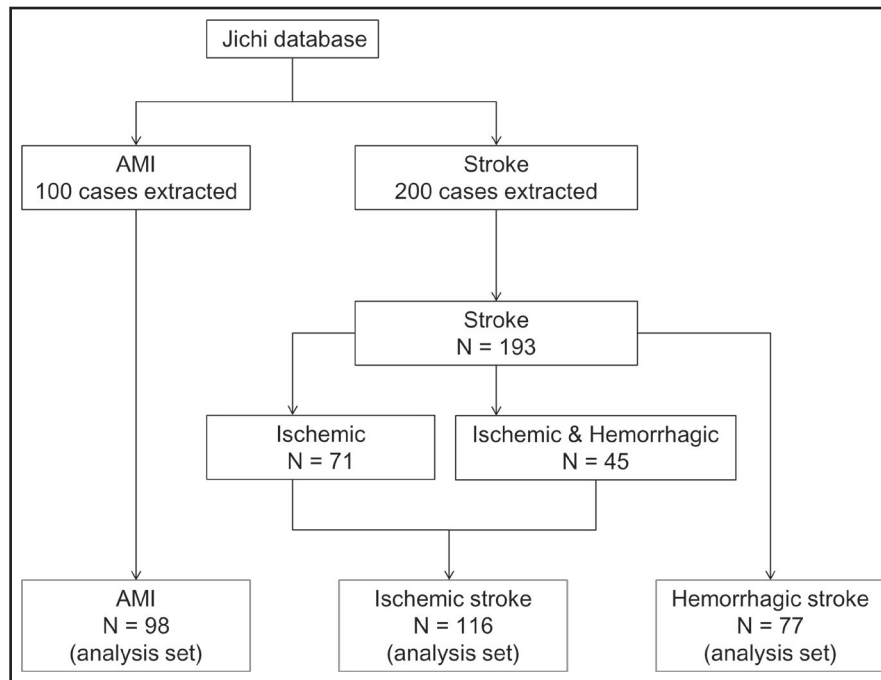


FIGURE 2 Study flow chart. AMI, acute myocardial infarction

limit (URL) and elevated cardiac troponin T or I with any of the following findings: (a) new electrocardiogram changes indicating ischemia (new ST-T changes or left bundle branch block [LBBB]), pathologic Q waves, or echocardiographic images of new cardiac regional wall motion abnormalities; (b) cardiovascular death associated with a new finding of ST elevation before an increase in biomarkers in the laboratory blood tests, with LBBB or new intracoronary thrombus confirmed by angiography or autopsy; (c) myocardial markers increased to ≥ 3 times the URL within 24 h after percutaneous coronary intervention; (d) myocardial markers increased to ≥ 5 times the URL and any of the following findings present within 24 h after coronary artery bypass grafting: new onset of Q wave (≥ 0.04 s) observed in 2 or more contiguous leads or LBBB, new occlusion of the graft or coronary artery confirmed by coronary angiography, or new imaging evidence of cardiac regional wall motion abnormalities; (e) Q waves (≥ 0.04 s) observed in 2 or more contiguous leads; and (f) pathological signs (eg, chest pain, sweating, and nausea) of acute MI.

A definitive diagnosis of stroke required a new acute onset of focal neurological symptoms lasting 24 h or more not caused by trauma or a known nonvascular condition (eg, brain tumor). Stroke events were classified into 4 types based on the standards of the American Heart Association/American Stroke Association using diagnostic imaging by CT or MRI or as described previously¹⁹: hemorrhagic, ischemic, infarction with subsequent hemorrhage, or unknown type. Hemorrhagic stroke was diagnosed by imaging evidence of intraparenchymal or subarachnoid hemorrhage or by lumbar puncture, neurosurgery, or autopsy. Ischemic stroke was defined by the presence of focal neurological disorders due to thrombi or emboli that partially remained 24 h after diagnosis. Infarction with subsequent hemorrhage was defined as an ischemic stroke in which

hemorrhage was initially absent and only observed during subsequent imaging. Lastly, unknown-type stroke was defined to include all cases without sufficient evidence to be classified as one of the above.

2.5 | Sample size

A set of 100 cases was initially planned to be extracted for acute MI, ischemic stroke, and hemorrhagic stroke from a feasibility perspective. However, we combined the 2 types of strokes for extraction (200 cases of stroke) because hemorrhagic stroke is a possible complication of ischemic stroke.

2.6 | Data analysis

Using the CEC classifications as the gold standard, we evaluated the validity of the potential event datasets using positive predictive values (PPVs) and their 95% confidence interval, the latter of which were calculated using the Clopper-Pearson method. The PPV was defined as the proportion of cases with incident cardiovascular events as judged by the CEC (ie, true cases) among the cases with potential cardiovascular events selected by the algorithms. Complementing this approach, we also calculated the apparent sensitivity, which we defined as the proportion of true cases selected by a given algorithm among all true cases identified by diagnosis only (100%). We assumed that cases identified by diagnosis only contained most of the true cases.²⁰ We performed the statistical analysis using SAS version 9.4 (SAS Institute, Inc).

3 | RESULTS

3.1 | Descriptive statistics

During the study period, we identified 540 acute MIs and 1173 all-cause stroke cases based on ICD-10 definitive diagnoses. Among these cases, we randomly selected 100 cases for the acute MI cohort and 200 cases for the stroke cohort. From the MI and stroke cohorts, 2 and 7 cases with no corresponding clinical data were excluded, respectively. Of the stroke set, 45 cases with ischemic and hemorrhagic stroke were included only in the ischemic stroke analysis set. The final analysis sets for MI, ischemic stroke, and hemorrhagic stroke included 98, 116, and 77 cases, respectively (Figure 2). Tables 1 and 2 summarize the characteristics, recorded diagnoses, prescribed drugs, and medical procedures of the included cohorts.

TABLE 1 Patient characteristics, claims-based diagnoses, drugs, and medical procedures of patients diagnosed with acute myocardial infarction

		Number of patients (%)
Total		98 (100)
Sex	Female	20 (20.4)
	Male	78 (79.6)
Age	Mean (SD)	66.8 (11.9)
Diagnoses	Acute inferior MI	69 (70.4)
	Acute posterior MI	14 (14.3)
	Acute inferior MI + Acute anterior MI	4 (4.1)
	Acute inferior MI + Acute MI	3 (3.1)
	Acute posterior MI + MI	2 (2.0)
	Others	6 (6.1)
Drugs	Aspirin + Clop	63 (64.3)
	Aspirin	11 (11.2)
	Aspirin + Warf + Clop	5 (5.1)
	Aspirin + Warf	4 (4.1)
	NOAC + Aspirin + Clop	2 (2.0)
	Others	7 (7.1)
	No drug	6 (6.1)
Medical procedures	CorAngio + Cardiac marker + Hosp	57 (58.2)
	CorAngio + Cardiac marker	24 (24.5)
	Cardiac marker	12 (12.2)
	Cardiac marker + Hosp	4 (4.1)
	No procedure	1 (1.0)

Abbreviations: Clop, clopidogrel sulfate; CorAngio, coronary angioplasty including percutaneous stent placement, percutaneous transluminal coronary angioplasty, and percutaneous thromboaspiration; Hosp, hospitalization; MI, myocardial infarction; NOAC, non-vitamin K antagonist oral anticoagulants; SD, standard deviation; Warf, warfarin potassium.

3.2 | Primary analysis

After the CEC assessed the potential cases, the PPVs were determined for each event based on the various extract algorithms (Table 3). The algorithms' PPVs differed considerably among cardiovascular events when considering the ICD-10 diagnosis only (acute MI, 81.6%; ischemic stroke, 31.0%; and hemorrhagic stroke, 45.5%). To remove true negative events, we examined algorithms consisting of combinations of parameters (eg, diagnoses + drugs, diagnoses + medical procedures, and diagnoses + drugs + medical procedures). The algorithms with the highest PPVs were diagnoses + drugs + medical procedures for acute MI (PPV, 87.0% each), diagnoses + drugs + medical procedures for ischemic stroke (PPV, 44.4%), and diagnosis + medical procedure for hemorrhagic stroke without ischemic stroke (PPV, 46.1%; Table 3). For acute MI, the simpler algorithm was preferred because the PPVs were similar.

3.3 | Exploratory analysis

Lastly, we conducted an exploratory analysis of the algorithms that employed subgroups of the factors defining the diagnoses, drugs, and medical procedures (Table 4). For acute MI, the potential event cases identified by diagnoses + drugs (clopidogrel sulfate) + medical procedures (hospitalization) and by diagnoses + drugs (clopidogrel sulfate) + medical procedures (hospitalization and coronary angioplasty) showed the highest PPVs (both 98.0%; 49/50 cases) and apparent sensitivities (both 61.3%). For ischemic stroke, the cases identified by diagnoses + drugs + medical procedures (ultrasonography and rehabilitation) and by diagnoses + drugs + medical procedures (ultrasonography, MRI/CT, and rehabilitation) showed the highest PPVs (both 70.0%; 21/30 cases) and apparent sensitivities (both 58.3%). For hemorrhagic stroke not including ischemic stroke, the potential event cases identified by diagnoses + medical procedures (MRI/CT and rehabilitation) and by diagnoses + medical procedures (rehabilitation) showed the highest PPVs (69.2% [27/39 cases] and 67.5% [27/40 cases], respectively) and apparent sensitivities (both 77.1%).

4 | DISCUSSION

In this validation study, we demonstrated that the PPVs of potential event cases are low when relying on ICD-10 codes alone and can be improved when extracted by algorithms using ICD-10 codes and other medical information (eg, prescription and procedure data). These novel algorithms might help improve the quality of the real-world evidence when identifying true incident cases of cardiovascular events in Japanese claims databases. It is possible to provide more appropriate procedures and treatment options in accordance with evidence-based clinical practices including the use of safer and more effective medicines.

TABLE 2 Patient characteristics and claims-based diagnoses, drugs, and medical procedures for patients diagnosed with stroke

		Number of patients (%)	
		Ischemic	Hemorrhagic (not ischemic)
Total		116 (100%)	77 (100%)
Sex	Female	50 (43.1%)	33 (42.9%)
	Male	66 (56.9%)	44 (57.1%)
Age	Mean (SD)	66.8 (13.1)	68.1 (13.9)
Diagnoses	Cerebral infarction	19 (16.4%)	—
	Cerebral hemorrhage + Cerebral infarction	13 (11.2%)	—
	Cerebellar infarction	8 (6.9%)	—
	Subarachnoid hemorrhage + Cerebral infarction	6 (5.2%)	—
	Cerebral hemorrhage	—	21 (27.3%)
	Subarachnoid hemorrhage	—	17 (22.1%)
	Chronic subdural hematoma	—	13 (16.9%)
	Putamen hemorrhage	—	6 (7.8%)
	Subcortical hemorrhage	—	6 (7.8%)
	Others	70 (60.3%)	14 (18.2%)
	Drugs	No Drug	39 (33.6%)
Warfarin potassium		12 (10.3%)	9 (11.7%)
Clopidogrel sulfate		10 (8.6%)	4 (5.2%)
Cilostazol		5 (4.3%)	5 (6.5%)
Aspirin		9 (7.8%)	—
Aspirin + warfarin potassium		3 (2.6%)	4 (5.2%)
Others		38 (32.8%)	18 (23.4%)
Medical procedures		MRI/CT	33 (28.4%)
	MRI/CT + Rehab + US	22 (19.0%)	9 (11.7%)
	MRI/CT + Rehab	4 (3.4%)	19 (24.7%)
	MRI/CT + Rehab + Hosp	12 (10.3%)	5 (6.5%)
	MRI/CT + Rehab + US + Hosp	11 (9.5%)	6 (7.8%)
	MRI/CT + US	12 (10.3%)	3 (3.9%)
	No procedure	11 (9.5%)	1 (1.3%)
	MRI/CT + Hosp	3 (2.6%)	5 (6.5%)
	MRI/CT + US + Hosp	4 (3.4%)	2 (2.6%)
	US	3 (2.6%)	—
	Rehab	1 (0.9%)	—
	Rehab + Hosp	—	1 (1.3%)

Abbreviations: CT, computed tomography; Hosp, hospitalization MRI, magnetic resonance imaging; Rehab, rehabilitation; SD, standard deviation; US, ultrasonography.

The accuracy of any algorithm for identifying target diseases varies greatly depending on the database employed and the target event.^{12,16} Indeed, our study observed a marked difference between acute MI and stroke in terms of this accuracy. In the primary analysis, we compared the PPVs among potential event cases identified by 4 algorithms. When we compared the PPVs of potential cases identified by ICD-10 diagnosis alone, stroke showed much lower PPVs

(ischemic stroke, 31.0%; hemorrhagic, 45.5%) compared with acute MI (PPV: 81.6%). Even the best algorithms provided the highest PPVs of 44.4%, 46.1%, and 87.0% for ischemic stroke, hemorrhagic stroke, and acute MI, respectively. The accuracy results illustrates that PPVs vary considerably across disease events even when using diagnoses with a “definitive” status in the claims databases. The low PPVs for stroke might be due to the high number of inconclusive

TABLE 3 Positive predictive values by algorithm based on the available claims data

	Claims-based algorithm	No. of potential cases	No. of true cases	PPV (%)	95% CI of PPV
Acute MI	Diagnoses only	98	80	81.6	(72.5, 88.7)
	Diagnoses + Drugs	92	80	87.0	(78.3, 93.1)
	Diagnoses + Medical procedures	97	80	82.5	(73.4, 89.4)
	Diagnoses + Drugs + Medical procedures	92	80	87.0	(78.3, 93.1)
Ischemic stroke	Diagnoses only	116	36	31.0	(22.8, 40.3)
	Diagnoses + Drugs	77	32	41.6	(30.4, 53.4)
	Diagnoses + Medical procedures	105	36	34.3	(25.3, 44.2)
	Diagnoses + Drugs + Medical procedures	72	32	44.4	(32.7, 56.6)
Hemorrhagic stroke (not ischemic)	Diagnoses only	77	35	45.5	(34.1, 57.2)
	Diagnoses + Medical procedures	76	35	46.1	(34.5, 57.9)

Note: Acute myocardial infarction (MI): diagnoses (I21 acute myocardial infarction, I22 recurrent myocardial infarction); drugs (anticoagulant, heparin, anti-thrombin, aspirin, thrombolytic agent); medical procedures (coronary angioplasty, myocardial marker tests, hospitalization). Ischemic stroke: diagnoses (I63 cerebral infarction); drugs (antiplatelet drug, anticoagulant, heparin, anti-thrombin, aspirin, cerebral metabolism drug, thrombolytic agent); medical procedures (head imaging [MRI/CT], carotid artery ultrasound, hospitalization, rehabilitation). Hemorrhagic stroke (not ischemic): diagnoses (I60 subarachnoid hemorrhage, I61 intracerebral hemorrhage, I62 non-traumatic intracranial hemorrhage); medical procedures (head imaging [MRI/CT], hospitalization, rehabilitation).

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging; PPV, positive predictive value.

diagnoses based on clinical findings, and the need for specific tests before a definitive diagnosis.²¹ Moreover, definitive diagnoses are often not recorded in claims databases, even after the final diagnosis has been reached through further clinical examinations, which would have lowered the associated PPVs. This is a common practice in many clinical institutions in Japan.^{21,22}

We hypothesized that careful examination of the factors affecting PPVs in each target disease could improve the PPV without loss of sensitivity. In the exploratory analysis of algorithms, which included the codes defining the drugs and medical procedures, we identified factors that increased the PPVs of cardiovascular events. Clopidogrel sulfate, hospitalization, and coronary angioplasty all increased the PPVs for acute MI in the Jichi database. However, for the routine use of the acute MI algorithm in general, "antiplatelet drugs" might be employed instead of "clopidogrel sulfate". Ultrasonography, MRI/CT, and rehabilitation increased the PPVs for ischemic stroke, whereas only MRI/CT and rehabilitation increased the PPVs for hemorrhagic stroke without ischemic stroke. Given that rehabilitation is critical for alleviating the sequelae of stroke, regardless of its severity,²³ rehabilitation-based algorithms are a plausible method for identifying true stroke events. Our data also showed that ultrasonography is a key factor in identifying true ischemic stroke events, which could be attributed to the establishment of mandatory procedures for detecting cardiac thrombi, with or without atrial fibrillation, in the routine management of ischemic stroke in Japan.²⁴ Overall, our results indicate that algorithms combining event-specific data on procedures, treatments, and post-treatment care could improve the validity of database research in true cardiovascular events.

High sensitivities, specificities, negative predictive values, and PPVs are all important measures when developing a reliable

algorithm, but there are tradeoffs depending on the aims of the research.²⁰ In this study, we prioritized achieving high PPVs for identifying cardiovascular events and did so without an apparent loss of sensitivity. Moreover, the exploratory analysis found more favorable algorithms with apparent sensitivity values of approximately 60%, which would allow for valid research of claims databases given that this sensitivity value would allow us to employ the National Database of Health Insurance Claims and Specific Health Checkups of Japan, a comprehensive database of health insurance claims data under Japan's National Health Insurance system, and enables retrospective cohort studies with a sample size of approximately 100 million with a very small selection bias.²⁵

Our study has certain limitations. First, there is no method for identifying acute MI or stroke that would not have been coded as such initially (in the ICD-10 classification), although physicians prioritize diseases by their life-threatening impact for the diagnosis, which could lead to the underreporting of events. Also, the ICD-11 codes have been developed. Although we determined the patient population with the ICD-10 code to identify the acute MI or stroke, there are no data that the disease code difference affects the result. Second, although the exploratory algorithms employing variables of specific drugs and medical procedures increased the PPVs, they also slightly reduced the apparent sensitivities. Third, using these variables could define populations of near-optimally managed patients with acute MI or stroke. By contrast, outcomes might be poorer for patients who are not managed optimally, which could jeopardize patients with the most severe forms of the events. Rehabilitation was employed for the stroke algorithm, which meant that it counts only the patients who survived the initial hospital stay but not the lethal events. Our proposed

TABLE 4 Positive predictive values by algorithm, based on the available claims data (exploratory analyses)

	Claims-based algorithm	No. of potential cases	No. of true cases (Apparent sensitivity ^a ; %)	PPV (%)
Acute MI ^d	Diagnosis ^b only	98	80 (100)	81.6
	Diagnoses ^b + Drugs (Clop) + MedPros (CorAngio)	74	70 (87.5)	94.6
	Diagnoses ^b + MedPros (Hosp + CorAngio)	57	55 (68.8)	96.5
	Diagnoses ^b + Drugs ^c + MedPro (Hosp + CorAngio)	57	55 (68.8)	96.5
	Diagnoses ^b + Drugs (Clop) + MedPro (Hosp)	50	49 (61.3)	98.0
	Diagnoses ^b + Drugs (Clop) + MedPros (Hosp + CorAngio)	50	49 (61.3)	98.0
Ischemic stroke ^d	Diagnosis ^b only	116	36 (100)	31.0
	Diagnoses ^b + Drugs ^c + MedPros (MRI/CT + Rehab)	43	26 (72.2)	60.5
	Diagnoses ^b + MedPros (US + Rehab)	33	22 (61.1)	66.7
	Diagnoses ^b + MedPros (US + MRI/CT + Rehab)	33	22 (61.1)	66.7
	Diagnoses ^b + Drugs ^c + MedPros (US + Rehab)	30	21 (58.3)	70.0
	Diagnoses ^b + Drugs ^c + MedPros (US + MRI/CT + Rehab)	30	21 (58.3)	70.0
Hemorrhagic stroke ^d (not ischemic)	Diagnosis ^b only	77	35 (100)	45.5
	Diagnoses ^b + MedPros (MRI/CT)	75	35 (100)	46.7
	Diagnoses ^b + MedPros (Rehab)	40	27 (77.1)	67.5
	Diagnoses ^b + MedPros (MRI/CT + Rehab)	39	27 (77.1)	69.2

Abbreviations: Clop, clopidogrel sulfate; CorAngio, coronary angioplasty, including percutaneous stent placement, percutaneous transluminal coronary angioplasty, and percutaneous thrombo-aspiration; CT, computed tomography; Hosp, hospitalization; MedPros, medical procedures; MRI, magnetic resonance imaging; PPV, positive predictive value; Rehab, rehabilitation; US, ultrasonography.

^aProportion of number of cases identified by the claims-based algorithm among all cases confirmed by the CEC from potential cases based on diagnosis only.

^bFull diagnosis combinations as shown in Table 3.

^cFull drug combinations as shown in Table 3.

^dTop 5 algorithms with apparent sensitivity >0.5 are shown for acute myocardial infarction and ischemic stroke. Top 3 algorithms with apparent sensitivity >0.5 are shown for hemorrhagic stroke (not ischemic).

algorithms might therefore be inappropriate for estimating the absolute incidence rates but might still be suitable for estimating the ratios of incidence rates between treatments. Fourth, given that we utilized the claims database of a single university hospital medical institution in Japan, our results may not be generalized to the databases of other medical institutions including general hospitals or other countries due to the application of different procedures and treatment options. However, stroke and acute MI are severe events associated with hospitalization during initial therapy in well-equipped medical institutions in Japan, and the parameters for the extraction algorithms in this study were consistent with the practice guidelines for stroke and acute MI. The system for medical fee processing in Japan²⁶ and the practice guidelines for treating cardiovascular events are standardized.^{24,27,28} Patients can therefore receive similar medical services regardless of wealth, and we can assume that there are no major differences in clinical practice among healthcare institutions in Japan. Lastly, despite

the random patient selection sampling employed in this study, the small sample size might have an effect on sampling variation.

As a last point, we would like to draw attention to the state of claims database validation in Japan. The newly revised regulation for Good Post-Marketing Study Practice opened the door to database research for post-marketing safety studies,²⁹ expanding the potential of database research in safety evaluations in Japan. To improve the quality of database research, it is crucial to use administrative data linkage for combining detailed individual-based information from multiple data sources, although infrastructures for data linkage are still insufficient in Japan.³⁰ If we consider the burden on medical institutions, as was experienced in our study, then the totality of the burden over multiple endpoints, drugs, and pharmaceutical companies seems insurmountable. A report on claims data validation has been produced by the Japanese Society of Pharmacoepidemiology that summarizes the current state of validation in Japan.³⁰ We recommend this report as a source for further information.

5 | CONCLUSION

Algorithms that rely on ICD-10 diagnostic codes in combination with data on specific drugs and medical procedures achieved PPVs exceeding 70% with reasonable apparent sensitivities and appear to be valid for identifying cardiovascular events in Japanese claims databases.

ACKNOWLEDGMENTS

We are grateful to WysiWyg Co., Ltd. for their assistance in editing the manuscript.

CONFLICT OF INTEREST

DS, YI, and SH are full-time employees at Pfizer Japan Inc KK has received research grants from A&D Co., Bayer Yakuhin, Boehringer Ingelheim, Daiichi Sankyo, EA Pharma, Fukuda Denshi, Medtronic, Mitsubishi Tanabe Pharma Corporation, Mochida Pharmaceutical Co., Omron Healthcare, Otsuka, Pfizer, Takeda, and Teijin Pharma; and honoraria from Daiichi Sankyo, Omron Healthcare, and Takeda. The other authors declared no conflicts of interest.

AUTHOR CONTRIBUTIONS

DS was responsible for the study's conception, design, and data interpretation, and the drafting and revision of the manuscript. YI, S.Hi., TK, S.Ho., K.Ko., SF, SN, NT, and K.Ka. contributed to reviewing the study's design, data interpretation, and the drafting and revision of the manuscript. TK performed the data extraction, and K.Ko., SF, and S.N. conducted an independent clinical event committee. All authors were also responsible for the decision to submit the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ORCID

Daisuke Shima  <https://orcid.org/0000-0003-2348-6210>

Shingo Higa  <https://orcid.org/0000-0002-9430-7809>

Satoshi Hoshide  <https://orcid.org/0000-0001-7541-5751>

Naoko Tomitani  <https://orcid.org/0000-0002-1443-7073>

Kazuomi Kario  <https://orcid.org/0000-0002-8251-4480>

REFERENCES

- World Health Organization. *Cardiovascular Diseases (CVDs)*. 2017. [http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Accessed August 7, 2020
- Townsend N, Wilson L, Bhatnagar P, et al. Cardiovascular disease in Europe: epidemiological update 2016. *Eur Heart J*. 2016;37:3232-3245.
- Ministry of Health, Labour and Welfare, Japan. IV. Analysis by Cause of Death. 2016. <https://www.mhlw.go.jp/english/database/db-hw/lifetb16/dl/lifetb16-04.pdf>. Accessed August 7, 2020
- Redon J, Tellez-Plaza M, Orozco-Beltran D, et al. Impact of hypertension on mortality and cardiovascular disease burden in patients with cardiovascular risk factors from a general practice setting: the ESCARVAL-risk study. *J Hypertens*. 2016;34:1075-1083.
- Ribeiro RA, Ziegelmann PK, Duncan BB, et al. Impact of statin dose on major cardiovascular events: a mixed treatment comparison meta-analysis involving more than 175,000 patients. *Int J Cardiol*. 2013;166:431-439.
- Ward S, Lloyd Jones M, Pandor A, et al. A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess*. 2007;11(14):1-160.
- Cutrona SL, Toh S, Iyer A, et al. Validation of acute myocardial infarction in the Food and Drug Administration's Mini-Sentinel program. *Pharmacoepidemiol Drug Saf*. 2013;22:40-54.
- Ishiguro C, Takeuchi Y, Uyama Y, et al. The MIHARI project: establishing a new framework for pharmacoepidemiological drug safety assessments by the Pharmaceuticals and Medical Devices Agency of Japan. *Pharmacoepidemiol Drug Saf*. 2016;25:854-859.
- Etminan M, Sodhi M, Samii A, et al. Tumor necrosis factor inhibitors and risk of peripheral neuropathy in patients with rheumatic diseases. *Semin Arthritis Rheum*. 2019;48:1083-1086.
- Liou J-T, Lin CW, Tsai C-L, et al. Risk of severe cardiovascular events from add-on tiotropium in chronic obstructive pulmonary disease. *Mayo Clin Proc*. 2018;93:1462-1473.
- Yasunaga H, Matsui H, Horiguchi H, et al. Clinical epidemiology and health services research using the diagnosis procedure combination database in Japan. *Asian Pac J Dis Manage*. 2013;7:19-24.
- U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for industry and FDA staff. Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. 2013. <https://www.fda.gov/oc/oads/drugs/guidances/ucm243537.pdf>. Accessed August 7, 2020
- Tyree PT, Lind BK, Lafferty WE. Challenges of using medical insurance claims data for utilization analysis. *Am J Med Qual*. 2006;21:269-275.
- World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th revision. <http://apps.who.int/classifications/icd10/browse/2016/en>. Accessed August 7, 2020
- Lanes SF, de Luise C. Bias due to false-positive diagnoses in an automated health insurance claims database. *Drug Saf*. 2006;29:1069-1075.
- Matsuda S, Fujimori K. The claim database in Japan. *Asian Pac J Dis Manage*. 2012;6:55-59.
- Tanaka S, Seto K, Kawakami K. Pharmacoepidemiology in Japan: medical databases and research achievements. *J Pharm Health Care Sci*. 2015;1:16.
- Thygesen K, Alpert JS, White HD. Universal definition of myocardial infarction. *J Am Coll Cardiol*. 2007;50:2173-2195.
- Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44:2064-2089.
- Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65:343-349.e2.
- Endo S, Ikeda N, Kondo T, et al. Development of an annually updated Japanese national clinical database for chest surgery in 2014. *Gen Thorac Cardiovasc Surg*. 2016;64:569-576.
- Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. *Biol Pharm Bull*. 2015;38:53-57.
- Winstein CJ, Stein J, Arena R, et al. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2016;47:e98-e169.
- Shinohara Y, Yanagihara T, Abe K, et al. II. Cerebral infarction/transient ischemic attack (TIA). *J Stroke Cerebrovasc Dis*. 2011;20:S31-73.
- Nakao YM, Miyamoto Y, Ueshima K, et al. Effectiveness of nationwide screening and lifestyle intervention for abdominal obesity and cardiometabolic risks in Japan: the metabolic syndrome and

- comprehensive lifestyle intervention study on nationwide database in Japan (MetS ACTION-J study). *PLoS One*. 2018;13:e0190862.
26. Reich MR, Ikegami N, Shibuya K, et al. 50 years of pursuing a healthy society in Japan. *Lancet*. 2011;378:1051-1053.
 27. Ishihara M, Fujino M, Ogawa H, et al. Clinical presentation, management, outcome of Japanese patients with acute myocardial infarction in the troponin era - Japanese registry of acute myocardial infarction diagnosed by universal definition (J-MINUET). *Circ J*. 2015;79:1255-1262.
 28. Shinohara Y, Yanagihara T, Abe K, et al. III. Intracerebral hemorrhage. *J Stroke Cerebrovasc Dis*. 2011;20:S74-99.
 29. Ministry of Health, Labour and Welfare, Japan. Revision of the Ministerial Ordinance on Good Postmarketing Study Practice (GPSP ordinance). Pharmaceuticals and Medical Devices Safety Information. No. 355 (August 2018). <https://www.pmda.go.jp/files/000225335.pdf> Accessed August 7, 2020
 30. Iwagami M, Aoki K, Akazawa M, et al. Task force report on the validation of diagnosis codes and other outcome definitions in the Japanese receipt data. *Jpn J Pharmacoepidemiol*. 2018;23:95-123 (in Japanese).

How to cite this article: Shima D, li Y, Higa S, et al. Validation of novel identification algorithms for major adverse cardiovascular events in a Japanese claims database. *J Clin Hypertens*. 2021;23:646–655. <https://doi.org/10.1111/jch.14151>