Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validation of Algorithms to Identify Acute Myocardial Infarction, Stroke, and Cardiovascular Death in German Health Insurance Data

Katharina Platzbecker[1], Annemarie Voss[1], Jonas Reinold[1], Anne Elbrecht[2], Wolfgang Biewener[2], Daniel Prieto-Alhambra[3,4], Annika M Jödicke[3], Tania Schink[1]

[1]Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany; [2]Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany; [3]Pharmaco- and Device Epidemiology, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

Correspondence: Tania Schink, Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Achterstrasse 30, Bremen, 28359, Germany, Tel +4942121856865, Email schink@leibniz-bips.de

Purpose: Validation of outcomes allows measurement of and correction for potential misclassification and targeted adjustment of algorithms for case definition. The purpose of our study was to validate algorithms for identifying cases of acute myocardial infarction (AMI), stroke, and cardiovascular (CV) death using patient profiles, ie, chronological tabular summaries of relevant available information on a patient, extracted from pseudonymized German claims data.

Patients and Methods: Based on the German Pharmacoepidemiological Research Database (GePaRD), 250 cases were randomly selected (50% males) for each outcome between 2016 and 2017 based on the inclusion criteria age ≥50 years and continuous insurance ≥1 year and applying the following algorithms: hospitalization with a main diagnosis of AMI (ICD-10-GM codes I21.- and I22.-) or stroke (I63, I61, I64) or death with a hospitalization in the 60 days before with a main diagnosis of CV disease. Patient profiles were built including (i) age and sex, (ii) hospitalizations incl. diagnoses, procedures, discharge reasons, (iii) outpatient diagnoses incl. diagnostic certainty, physician specialty, (iv) outpatient encounters, and (v) outpatient dispensings. Using adjudication criteria based on clinical guidelines and risk factors, two trained physicians independently classified cases as "certain", "probable", "unlikely" or "not assessable". Positive predictive values (PPVs) were calculated as percentage of confirmed cases among all assessable cases.

Results: For AMI, the overall PPV was 97.6% [95% confidence interval 94.8–99.1]. The PPV for any stroke was 94.8% [91.3–97.2] and higher for ischemic (98.3% [95.0–99.6]) than for hemorrhagic stroke (86.5% [76.5–93.3]). The PPV for CV death was 79.9% [74.4–84.4]. It increased to 91.7% [87.2–95.0] after excluding 32 cases with data insufficient for a decision.

Conclusion: Algorithms based on hospital diagnoses can identify AMI, stroke, and CV death from German claims data with high PPV. This was the first study to show that German claims data contain information suitable for outcome validation.

Keywords: claims data, algorithm validation, patient profiles, positive predictive value

Plain Language Summary

Correct assessment of study endpoints is essential in research on drug safety. We aimed to validate algorithms for the identification of the endpoints acute myocardial infarction (AMI), stroke, and cardiovascular (CV) death in German claims data using patient profiles, ie, chronological tabular summaries of the relevant information available on a patient, extracted from de-identified German claims data. For each of the three endpoints, we sampled 250 cases, 125 in women and 125 in men, from the German Pharmacoepidemiological Research Database (GePaRD) applying our algorithms for AMI, stroke, and CV death. For each sampled case, we created a de-identified patient profile including age and sex, information on hospitalizations, outpatient diagnoses incl. diagnostic certainty and physician specialty, outpatient encounters, and outpatient medication dispensings. Two trained physicians independently reviewed all profiles. For AMI and stroke, 97.6% and 94.8% of the cases were confirmed. For CV death, 79.9% were confirmed. After the exclusion of 32 cases with data insufficient for a decision, the proportion of confirmed cases increased to 91.7%.

Received: 28 June 2022
Accepted: 30 September 2022
Published: 10 November 2022

In conclusion, the endpoints AMI, stroke, and CV death can be validly assessed in German claims data. We were also the first to show that German claims data contain information suitable to validate study endpoints.

Introduction

Correct assessment of outcomes is essential in pharmacoepidemiological studies on the safety of medication. Validation of outcomes allows measurement of and correction for potential misclassification.[1–4] Furthermore, algorithms developed for case definition may be modified based on validation results.

In claims databases, linkage to patients' charts—which are usually the reference ("gold standard") in validation studies—is often impossible due to data privacy regulations. Health insurance data, however, include comprehensive, prospectively collected information on diagnoses, treatments, hospitalizations, and other interactions with the healthcare system, usually over a long period of time. Patient profiles extracted from French claims data, ie, chronological tabular summaries of the relevant information available on a patient, have been shown to serve as an alternative to patient charts.[5] To date, no study has been published on using patient profiles extracted from German claims data for outcome validation.

Cardiovascular (CV) outcomes, such as acute myocardial infarction (AMI), stroke, and CV death are—either individually or as the composite endpoint 'major cardiovascular events' (MACE)—potential major adverse drug reactions to be considered in drug safety studies informing regulators and treatment guidelines. Due to the impact these drug safety studies might have on clinical practice, correct assessment of outcomes is essential. Validation of outcome assessment is needed to increase the confidence of regulators and clinicians in the validity of results.

Several algorithms for the identification of the three CV outcomes under study have been published and validated.[6–9] However, the specification of algorithms depends on the information and level of detail available in the respective database and their performance differs between data sources. To our knowledge, to date, no algorithms for AMI, stroke, and CV death have been validated for German claims data.

The objective of our study was to assess the validity of algorithms to identify AMI, stroke, and CV death in German claims data using patient profiles, ie, chronological tabular summaries of the relevant information available on a patient extracted from pseudonymized German health insurance data.

Methods

Data Source

The German Pharmacoepidemiological Research Database (GePaRD) is based on claims data from four statutory health insurance providers in Germany and currently includes information on approximately 25 million persons who have been insured with one of the participating providers since 2004 or later. Per data year, there is information on approximately 20% of the general population and all geographical regions of Germany are represented.

GePaRD contains demographic information such as year of birth, sex, and region of residence as well as information on hospitalizations, outpatient visits, and outpatient drug dispensings. Data on hospitalizations include the date of admission, the admission diagnosis, diagnostic and surgical/medical procedures during the hospital stay, the discharge date, main and secondary discharge diagnoses, and the reason for discharge (incl. death). Outpatient data include diagnoses as well as outpatient diagnostic and therapeutic procedures and services. Once per quarter, physicians in the outpatient setting code the disease(s) for which they treated their patients. Coding the diagnostic certainty is mandatory in the outpatient setting. This coding differentiates between "confirmed", "suspected", "status post", and "ruled out" diagnoses. Hospital and outpatient diagnoses are coded using the International Classification of Diseases, version 10 in the German Modification (ICD-10-GM) with at least four digits; diagnostic and surgical/medical procedures are coded using the Operations and Procedures Coding System (OPS), and outpatient treatment/diagnostic procedures as well as immunizations are coded using claim codes for outpatient services and procedures (German: Einheitlicher Bewertungsmaßstab, EBM).

GePaRD contains information on all physician-prescribed medication dispensed in a pharmacy and reimbursed by the health insurance provider. Information on medication is coded based on the German modification of the Anatomical

Therapeutic Chemical (ATC) Classification System. Information on medication purchased over the counter (OTC) is not available in GePaRD. With a few exceptions regarding expensive drugs (eg, monoclonal antibodies), there is no information on medication administered in the hospital.

For lab tests and physical exams, related information including the date is available in the database provided they are reimbursable. Results of these procedures are unavailable but can partly be derived indirectly if specific ICD-10-GM diagnoses or treatments are coded subsequently to tests or exams. There is no lifestyle information in GePaRD. Certain subgroups that have developed diseases due to an unhealthy lifestyle may be identified through diagnosis codes (eg, obesity, liver diseases due to alcohol abuse) or specific treatments.

In Germany, the utilization of health insurance data for scientific research is regulated by the Code of Social Law. All involved health insurance providers as well as the German Federal Insurance Office and the Senator for Science, Health, and Consumer Protection in Bremen as their responsible authorities approved the use of GePaRD data for this study. Informed consent for studies based on claims data is required by law unless obtaining consent appears unacceptable and would bias results, which was the case in this study. According to the Ethics Committee of the University of Bremen studies based on GePaRD are exempt from institutional review board review.

Outcome Definition, Sampling of Cases and Creation of Patient Profiles

An AMI event was defined as hospitalization with a main diagnosis of AMI (ICD-10-codes I21.- and I22.-). Accordingly, a stroke event was defined as hospitalization with a main diagnosis of stroke and classified as ischemic (I63.-), hemorrhagic (I61.-) or unspecified (I64). CV death was defined as a combination of death[10] and, within 60 days before the death date, a hospitalization with a main diagnosis of sudden cardiac death (I46.1, I46.9), heart failure (I11.0, I13.0, I13.2, I50.-), cardiac arrhythmia (I44.-, I47.-, I48., I49.-), stroke (I61.-, I63.-, I64), cerebrovascular disease (G45.3. G45.8, G45.9) or AMI (I21.-, I22.-).

All events in GePaRD between 1 January 2016 and 31 December 2017 were eligible for validation, if the respective patients fulfilled all of the following inclusion criteria: (i) 50 years or older at date of event, (ii) continuous insurance period of at least one year before date of the event, (iii) valid information on sex, and (iv) residency in Germany. Each occurring event was included separately if all inclusion criteria were fulfilled. Each patient could contribute one event per outcome. For each of the three outcomes, we randomly sampled 250 cases, thereof 125 each in females and males. For the outcome stroke, a distribution of 70% ischemic or unspecified and 30% hemorrhagic events was required in each sex stratum.

For each of the 750 cases, we retrieved (i) age and sex, (ii) all hospitalizations with length of stay, admission and discharge dates, all inpatient diagnoses, inpatient procedures, and discharge reason from up to five years before and up to one year after the event, (iii) all outpatient diagnoses including diagnostic certainty and specialty of the diagnosing physician from up to five years before and up to one year after the event, (iv) all outpatient encounters and procedures up to one year before and after the event, and (v) all outpatient dispensings up to one year before and after the event. To create the patient profiles, the information described above was compiled and transferred into tabular format. For each case, one table was created including all information—one row per claim—of the respective patient in chronological order. In addition to codes and descriptions, the timing in relation to the date of the event (day 0) was included (Table 1).

Review of Patient Profiles

Two physicians independently reviewed each patient profile. Based on the review and the experts' clinical judgement, cases were classified as "certain", "probable", "unlikely", or "not assessable".

In advance, adjudication criteria were compiled based on guidelines on diagnosis and treatment as well as risk factors of the respective outcome[11–15] and discussed with the reviewing physicians. Based on those criteria, ACCESS assessment forms were created to support the review process.

All reviewing physicians were trained with regard to the structure and content of GePaRD and the utilization of the assessment forms. The reviewers were not instructed on specific criteria for the classification of cases but were asked to apply their clinical judgement. With regard to the assessment scale, physicians classified a case as "certain" if the data available on the clinical course of the case clearly confirmed the diagnosis. "Probable" cases lacked data one would

Table 1 Excerpt from a Fictitious Exemplary Patient Profile for AMI Validation

ID	Age	Sex	Days from Event	Source	Code	Description of Code (German Modification)	Setting	Diagnosis Type	Physician Specification
1	75	F	−85	ICD	E789	Disorder of lipoprotein metabolism, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	I109	Essential hypertension, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	E049	Nontoxic goiter, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	E039	Hypothyroidism, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	G473	Sleep apnea	Outpatient	Certain	Primary care
1	75	F	−85	ICD	E669	Obesity, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	G560	Carpal tunnel syndrome	Outpatient	Suspected	Primary care
1	75	F	−85	ICD	M179	Osteoarthritis of knee, unspecified	Outpatient	Certain	Primary care
1	75	F	−85	ICD	M545	Low back pain	Outpatient	Certain	Primary care
1	75	F	−85	ATC	C10AA01	Simvastatin	Outpatient		Primary care
1	75	F	−85	ATC	A02BC02	Pantoprazole	Outpatient		Primary care
1	75	F	−85	ATC	C03CA04	Torasemide	Outpatient		Primary care
1	75	F	−85	ATC	H03AA01	Levothyroxine-Sodium	Outpatient		Primary care
1	75	F	−70	ICD	H333	Retinal break without retinal detachment	Outpatient	Status post	Ophthalmology
1	75	F	−70	ICD	H521	Myopia	Outpatient	Certain	Ophthalmology
1	75	F	−70	ICD	H522	Astigmatism	Outpatient	Certain	Ophthalmology
1	75	F	−70	ICD	H524	Presbyopia	Outpatient	Certain	Ophthalmology
1	75	F	−70	ICD	H310	Chorioretinal scars	Outpatient	Certain	Ophthalmology
1	75	F	−70	ICD	H269	Unspecified cataract	Outpatient	Certain	Ophthalmology
1	75	F	−70	ICD	Z961	Presence of intraocular lens implant	Outpatient	Certain	Ophthalmology
1	75	F	0	EBM	32150	Immunological detection of troponin I and/or troponin T	Outpatient		Primary care
1	75	F	0			Hospital Admission	Inpatient		
1	75	F	0	ICD	I100	Benign essential hypertension	Inpatient	Secondary diagnosis	
1	75	F	0	ICD	I251	Atherosclerotic heart disease	Inpatient	Secondary diagnosis	
1	75	F	0	ICD	E782	Mixed hyperlipidemia	Inpatient	Secondary diagnosis	
1	**75**	**F**	**0**	**ICD**	**I214**	**Acute subendocardial myocardial infarction***	**Inpatient**	**Main admission diagnosis, main discharge diagnosis**	
1	75	F	0	OPS	12752	Transarterial left heart catheter examination: coronary angiography, pressure measurement and ventriculography in the left ventricle	Inpatient		

(*Continued*)

Table 1 (Continued).

ID	Age	Sex	Days from Event	Source	Code	Description of Code (German Modification)	Setting	Diagnosis Type	Physician Specification
1	75	F	0	OPS	6002k0	Application of drugs, list 2: Eptifibatid, parenterally: 30 mg up to under 75 mg	Inpatient		
1	75	F	0	OPS	883700	Percutaneous transluminal vascular intervention on heart and coronary vessels: Balloon angioplasty: Single coronary artery	Inpatient		
1	75	F	0	OPS	8837m1	Percutaneous transluminal vascular intervention on heart and coronary vessels: Placement of a drug-eluting stent: 2 stents in one coronary artery	Inpatient		
1	75	F	0	OPS	883b0c	Additional information on materials: Everolimus-eluting stents or OPD systems with other polymer	Inpatient		
1	75	F	0	OPS	883b50	Additional information on materials: Use of a modelling or double lumen balloon: 1 modelling balloon	Inpatient		
1	75	F	0	OPS	883bc5	Additional information on materials: Use of a vascular occlusion system: resorbable plugs without anchor	Inpatient		
1	75	F	0	OPS	8930	Monitoring of respiration, heart and circulation without measuring pulmonary artery pressure and central venous pressure	Inpatient		
1	75	F	7			Hospital discharge, regular termination of treatment			
1	75	F	8	ATC	B01AC06	Acetylsalicylic acid	Outpatient		Primary care
1	75	F	8	ATC	B01AC24	Ticagrelor	Outpatient		Primary care
1	75	F	8	ATC	C07AB07	Bisoprolol	Outpatient		Primary care
1	75	F	8	ATC	C10AA05	Atorvastatin	Outpatient		Primary care
1	75	F	8	ATC	C09AA05	Ramipril	Outpatient		Primary care
1	75	F	54	ICD	G560	Carpal tunnel syndrome	Outpatient	Certain	Neurology
1	75	F	54	EBM	27311	Clinical neurological basic diagnostics	Outpatient		Neurology
1	75	F	54	EBM	27331	Evaluation of a peripheral neuromuscular disease	Outpatient		Neurology
1	75	F	57	ICD	I2520	Old myocardial infarction: past 29 days to 4 months	Outpatient	Certain	Primary care
1	75	F	57	ICD	I2519	Atherosclerotic heart disease, unspecified	Outpatient	Certain	Primary care

(*Continued*)

Table 1 (Continued).

ID	Age	Sex	Days from Event	Source	Code	Description of Code (German Modification)	Setting	Diagnosis Type	Physician Specification
1	75	F	57	ICD	E789	Disorder of lipoprotein metabolism, unspecified	Outpatient	Certain	Primary care
1	75	F	57	ICD	I109	Essential hypertension, unspecified	Outpatient	Certain	Primary care
1	75	F	57	ICD	E039	Hypothyroidism, unspecified	Outpatient	Certain	Primary care
1	75	F	57	ICD	G473	Sleep apnea	Outpatient	Certain	Primary care
1	75	F	57	ICD	E669	Obesity, unspecified	Outpatient	Certain	Primary care
1	75	F	57	EBM	3230	Problem-oriented medical consultation, which is necessary due to the nature and severity of the illness	Outpatient		Primary care
1	75	F	57	ATC	A02BC02	Pantoprazole	Outpatient		Primary care
1	75	F	57	ATC	H03AA01	Levothyroxine-Sodium	Outpatient		Primary care

Note: *Bold data indicate the inpatient discharge diagnosis defining the case.

expect in the context of a case (eg, specific diagnostic procedures needed to confirm a diagnosis, treatment, follow-up diagnoses). However, physicians still considered such cases more likely than a different outcome. In contrast, a different outcome was considered more likely in "unlikely" cases. For cases classified as "not assessable" the data were incomplete, preventing the physicians from providing a classification. In an initial training session, each physician reviewed 15 training patient profiles, followed by a discussion of remaining questions and the subsequent finalization of the assessment forms.

An additional category, "CV death cannot be excluded", was included as a result of those preparatory procedures. It captured cases for which differentiation between "probable" and "unlikely" was impossible with the provided information, eg, because potential cardiovascular and competing causes of death were considered to be equally likely.

Statistical Analysis

An event was considered confirmed if both reviewers categorized it as "definite" or "probable". Cases where reviewers disagreed led to a consensus conference with a third physician.

Positive predictive values (PPVs) were calculated as the percentage of confirmed cases ("certain" or "probable") among all cases of the respective outcome. Cases classified as "not assessable" were excluded from the analyses. Results were provided overall and stratified by sex.

Results

Acute Myocardial Infarction

A total of 70,818 AMI cases were observed in GePaRD between 1 January 2016 and 31 December 2017. Of these, 5520 cases were excluded overall because the patient was younger than 50 years (n = 4546), lacked a continuous insurance period of one year before the event (n = 1138), lacked valid information on sex (n = 8), and/or had no residence in Germany (n = 118, multiple exclusion criteria in one individual possible). Thus, 65,298 cases were eligible, thereof 22,804 cases in 19,758 women and 42,494 cases in 36,724 men (multiple cases in one individual possible).

Among the 250 sampled cases with a female to male ratio of 1:1, median [interquartile range (IQR)] age at the time of the event was 75 years [67–82]. Cardiovascular risk factors and comorbidities were common. For example, 71.2% of cases had dyslipidemia, 86.6% hypertension, 44.0% arrhythmia, 43.2% type 2 diabetes, and 38.0% heart failure. In line

Table 2 Agreement Between Reviewers Regarding AMI* Case Classification

		Reviewer 2			
		Definite/Probable	**Unlikely**	**Not Assessable**	
Reviewer1	**Definite/probable**	**228**[$]	0	8	236
	Unlikely	7	**4**	1	12
	Not assessable	0	0	**2**	2
		235	4	11	250

Notes: *Acute myocardial infarction. [$]Bold text indicates numbers of cases where both reviewers agreed.

with this observation, considerable proportions of cases had a history of beta-blocking (49.2%), lipid lowering (39.6%) or antidiabetic (20.8%) treatment (see Supplementary Table 1).

Agreement between reviewers was good. Both reviewers classified 93.6% of cases in the same category: 228 as "definite/probable", four as "unlikely" and two as "not assessable" (Table 2).

After discussion of the remaining 16 profiles in the consensus conference, 242 cases were classified as "definite/probable", six as "unlikely", and two as "not assessable", resulting in a PPV of 97.6% [95% confidence interval 94.8–99.1]. PPV in men was higher than in women (Table 3).

Table 3 Results of Validation of AMI*, Stroke and CV[$] Death Cases

Outcome	**Sample**	**Sample size**	**Case classification**				**PPV[#] (95% CI[§])**
			Definite/ probable	**Unlikely**	**Cannot be excluded**	**Not assessable**	
Acute myocardial infarction	**Total**	**250**	**242**	**6**		**2**	**97.6% (94.8–99.1)**
	Stratified by sex						
	Women	125	119	5		1	96.0% (90.8–98.7)
	Men	125	123	1		1	99.2% (95.6–100.0)
Stroke	**Total**	**250**	**237**	**13**		**0**	**94.8% (91.3–97.2)**
	Stratified by sex						
	Women	125	119	6		0	95.2% (89.8–98.2)
	Men	125	118	7		0	94.4% (88.8–97.7)
	Stratified by stroke type						
	Ischemic	173	170	3		0	98.3% (95.0–99.6)
	Hemorrhagic	74	64	10		0	86.5% (76.5–93.3)
	Unspecified	3	3	0		0	100% (29.2–100)
Cardio-vascular death	**Total**	**250**	**199**	**18**	**32**	**1**	**79.9% (74.4–84.7)**
	Stratified by sex						
	Women	125	101	10	13	1	81.5% (73.5–87.9)
	Men	125	98	8	19	0	78.4% (70.2–85.3)

Notes: *Acute myocardial infarction. [$]Cardiovascular. [#]Positive predictive value. [§]Confidence interval.

Stroke

Overall, 101,555 cases of stroke were observed in GePaRD between 1 January 2016 and 31 December 2017. Of those, 6477 were excluded for age <50 years (n = 5424), continuous insurance of <1 year (n = 1249), no valid information on sex (n = 5), and/or a place of residence outside of Germany (n = 156, multiple exclusion criteria may apply). This resulted in a total of 95,078 cases for sampling, including 46,803 cases among 38,471 women and 48,275 cases among 39,484 men.

Among the final sample of 250 cases of stroke with a female to male ratio of 1:1, median [IQR] age was 77 [68–78] years. Risk factors for stroke, such as hypertension (86.4%), arrhythmia (49.2%), dyslipidemia (67.2%) or type 2 diabetes (32.4%) were common. Overall, 43.2% of patients had a history of stroke, 8.4% of transient ischemic attack, and 9.2% suffered from other cerebrovascular diseases. Accordingly, betablockers (48.4%), lipid-lowering drugs (33.2%), antidiabetics (14.8%), and low-dose aspirin (13.2%) ranked among the most common medications (see Supplementary Table 2).

Agreement between reviewers was good. Both reviewers classified 95.6% of stroke cases in the same category: 231 as "definite/probable", eight as "unlikely" and none as "not assessable" (Table 4).

After discussing the remaining 11 profiles with the third reviewer, 237 were classified as "definite/probable", 13 as "unlikely" and none as "not assessable" yielding an overall PPV of 94.8% [91.3–97.2]. The PPV was comparable between women and men, but higher in cases of ischemic (98.3% [95.0–99.6]) compared to hemorrhagic stroke (86.5% [76.5–93.3], Table 3).

CV Death

A total of 37,604 cases of CV death were identified in GePaRD between 1 January 2016 and 31 December 2017. After exclusion of cases below the age of 50 years (n = 561), with a continuous insurance period of less than one year (n = 233), without valid information on sex (n = 1), and without a place of residence in Germany (n = 20), 36,833 eligible cases remained for sampling (multiple exclusion criteria may apply). Of those, 19,069 events occurred in women and 17,764 in men.

Among the final sample of 250 cases of CV death with a female to male ratio of 1:1, median [IQR] age was 82 [76–89] years. The burden of cardiovascular comorbidities and risk factors was high. For example, 79.2% had a history of heart failure, 76.8% of cardiac arrhythmia, and 50.8% of stroke (see Supplementary Table 3).

Agreement between reviewers was lower than for AMI and stroke. Both reviewers classified a total of 74.8% of cases of CV death in the same category: 162 as "definite/probable", 20 as "CV death cannot be excluded", five as "unlikely", and none as "not assessable" (Table 5).

After discussing the remaining cases with the third reviewer, 199 cases were classified as "definite/probable", 32 as "CV death cannot be excluded", 18 as "unlikely" and one as "not assessable", reaching a PPV of 79.9% [74.4–84.7] (Table 3). Excluding cases classified as "CV death cannot be excluded" in addition to "not assessable" cases from the denominator resulted in a PPV of 91.7% [86.5–94.6].

Most deaths (n = 209, 83.6%) were observed in hospital; 41 occurred in the outpatient setting. In the hospital setting, the PPV was higher (84.6% [79.0–89.2]) than in the outpatient setting (56.1% [39.8–71.5]). After exclusion of cases

Table 4 Agreement Between Reviewers Regarding Stroke Case Classification

		Reviewer 2			
		Definite/ probable	**Unlikely**	**Not assessable**	
Reviewer1	**Definite/probable**	**231***	2	4	237
	Unlikely	2	**8**	3	13
	Not assessable	0	0	**0**	0
		233	10	7	250

Note: *Bold text indicates numbers of cases where both reviewers agreed.

Table 5 Agreement Between Reviewers Regarding CV* Death Case Classification

		Reviewer 2				
		Definite/ probable	**Cannot be excluded**	**Unlikely**	**Not assessable**	
Reviewer1	**Definite/probable**	**162[$]**	6	2	0	170
	Cannot be excluded	28	**20**	19	3	70
	Unlikely	3	0	**5**	0	8
	Not assessable	2	0	0	**0**	2
		195	26	26	3	250

Notes: *Cardiovascular. [$]Bold text indicates numbers of cases where both reviewers agreed.

classified as "CV death cannot be excluded", the PPVs increased to 92.6% [88.0–95.9] and 85.2% [66.3–95.8], respectively.

Discussion

This validation study confirmed the high validity of hospital diagnoses for the ascertainment of AMI, stroke, and CV death in German claims data. In addition, we demonstrated that German claims data contain information for a comprehensive assessment of disease courses for outcome validation. Overall, reviewing physicians were satisfied with the level of detail available from the patient profiles.

Comparison of our PPVs for AMI and stroke algorithms with results from other studies is hampered by different coding-systems (eg, ICD-9), different settings (eg, inclusion of outpatient diagnoses), different gold standards, and—especially for stroke—different case definitions (eg, inclusion of subarachnoid hemorrhage). A meta-analysis by McCormick et al on AMI stated that the PPV was ≥89% in all eleven European studies and ≥93% in the two studies reporting the PPV separately for ICD-10-codes.[7] As in our study, the PPVs were higher in males than in females. Regarding stroke, McCormick et al reported in another meta-analysis a PPV of ischemic stroke (ICD-9 code 434, ICD-10 code I63) of ≥82% in 20 of the 27 studies reporting PPVs.[8] If unspecified strokes were included (433/434/436 and I63/I64), the PPV ranged from 46% to 94% in the 19 available studies. The PPV of hemorrhagic stroke (431, I61) was reported in 16 of the 25 studies with PPVs >87%. We found no published study validating CV death as defined in our study. Lix et al used a pre-existing cohort of patients receiving antidiabetic medication for their validation and opted for a much broader definition of CV death including, amongst others, death due to hypertensive disease.[16] The overall PPV was 54.5%, ranging from 34% to 73% for individual Canadian provinces. A meta-analysis by Singh et al included five studies, four of which reported PPVs for sudden cardiac death and one for AMI- and stroke-related death.[9]

While PPVs for AMI and stroke reached excellent values in our study, the PPV for the outcome CV death was lower. For the validation of this endpoint, reviewers did not only have to confirm the occurrence of the endpoint death, they also had to evaluate its etiology. Reviewers reported this evaluation of the specific cause of death to be most challenging if only a limited number of codes were available around the time of death. While the density of codes is usually high during hospitalizations due to close patient monitoring and documentation, fewer codes were observed in cases occurring outside the hospital setting. In addition, inpatient diagnosis types (eg, main discharge diagnosis, secondary diagnosis, admission diagnosis, etc.) facilitate conclusions on the severity and current relevance of diagnoses, while this differentiation of diagnosis types is impossible for outpatient diagnoses. In our study, most deaths (83.6%, n = 209) were observed in hospital, only 16.4% (n = 41) occurred in the outpatient setting. A total of 34.1% (n = 14) outpatient cases were classified as "CV death cannot be excluded". In contrast, only 8.6% (n = 18) inpatient cases were classified as "CV death cannot be excluded" and one (0.5%) as "not assessable". When reviewing all 32 cases classified as "CV death cannot be excluded", we found that in the majority of inpatient cases, the main discharge diagnosis of the hospitalization ending with death was not CV. Observed main discharge diagnoses included pneumonia in five cases, kidney failure in two cases, sepsis in two cases, and gastrointestinal diseases in two cases. Among the

14 outpatient cases classified as "CV death cannot be excluded", the mean time from the last hospitalization to death was 17 days (range: 2 to 40 days) and in four of the cases (28.6%), the last hospitalization had no CV main discharge diagnosis. In response to these observations, we conducted post-hoc analyses and found that the PPV was considerably higher when restricting the validation to patients who deceased in the hospital (84.6%). After excluding cases classified as CV death cannot be excluded', the PPVs increased to 92.6% among inpatient cases and to 85.2% among outpatient cases. Our observations are in line with previously published findings. A study based on administrative health records from Canada reported an increase in the PPV if restricting the algorithm of CV death to inpatient deaths with a CV main discharge diagnosis.[16]

This study is subject to limitations that need to be considered when interpreting the results. While claims data contain rich information on diagnoses, procedures, and health care received by patients, they lack clinical information like lab values and other diagnostic results. Also, death certificates are unavailable. In addition, the validation of outcome algorithms was performed with the same database that was used to develop the algorithms. Patient charts usually contain very detailed and accurate information and are therefore used as a reference ("gold standard") for case validation, ie, the diagnosis based on review of patients' charts is assumed to be correct. Chart validation is impossible in our database as due to strict data protection rules only pseudonymized data are available with regard to both the patients and the hospital or clinical practice. However, Thurin et al conclude that in the absence of an alternative, tabular patient profiles based on claims data ("reconstituted electronic health records") are a valuable tool for intra-database validation and performance estimation of case-identifying algorithms.[5] In fact, Thurin et al suggested specific conditions necessary for a successful validation process, which were satisfied in our study: "1) the health outcome of interest must be managed by a specific sequence of cares and encounters; and 2) the considered health-care database must capture in an exhaustive way a sufficient number of medical elements in line with the outcome of interest". Misclassification of outcomes might seriously bias study findings. Information on the validity of outcome assessment is essential for stakeholders such as regulators or clinicians as it enables them to evaluate drug safety studies and to determine potential consequences. Our study shows that German claims data can be used to validly assess CV safety endpoints.

Conclusion

Algorithms based on hospital diagnoses can identify AMI, stroke, and CV death from German claims data with high PPV. This was also the first study to show that German claims data contain information suitable for outcome validation.

Acknowledgments

We thank the statutory health insurance providers AOK Bremen/Bremerhaven, DAK-Gesundheit, Die Techniker (TK), and hkk Krankenkasse for contributing the data for this analysis. We would also like to thank Alina Ludewig, Marieke Niemeyer, and Philipp Alexander Volkmar for programming the analysis datasets and our statistician Malte Braitmaier.

Funding

This study was funded by UCB Biopharma UCL in the context of a post-authorization safety study (PASS) requested by the European Medicines Agency. The study was performed in line with the ENCePP Code of Conduct. The authors had complete autonomy in the process of establishing the protocol, carrying out the analyses, and interpreting the results and retained the full right to publish the results without limitation.

Disclosure

KP, AV, JR, AE, WB, and TS are working at an independent, non-profit research institute, the Leibniz Institute for Prevention Research and Epidemiology – BIPS. Unrelated to this study, BIPS occasionally conducts post-authorization safety studies (PASS) requested by health authorities which are financed by the pharmaceutical industry and performed in line with the ENCePP Code of Conduct. DPA's department has received grant/s from Amgen, Chiesi-Taylor, Lilly, Janssen, Novartis, and UCB Biopharma. His research group has received consultancy fees from Astra Zeneca and UCB Biopharma. Amgen, Astellas, Janssen, Synapse Management Partners and UCB Biopharma have funded or supported training programmes organised by DPA's department. The authors report no other conflicts of interest in this work.

References

1. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Statistics for Biology and Health. Springer; 2009.
2. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am J Epidemiol*. 1993;138(11):1007–1015. doi:10.1093/oxfordjournals.aje.a116805
3. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep*. 2014;1(4):175–185. doi:10.1007/s40471-014-0027-z
4. Hall GC, Lanes S, Bollaerts K, Zhou X, Ferreira G, Gini R. Outcome misclassification: impact, usual practice in pharmacoepidemiology database studies and an online aid to correct biased estimates of risk ratio or cumulative incidence. *Pharmacoepidemiol Drug Saf*. 2020;29(11):1450–1455. doi:10.1002/pds.5109
5. Thurin NH, Bosco-Levy P, Blin P, et al. Intra-database validation of case-identifying algorithms using reconstituted electronic health records from healthcare claims data. *BMC Med Res Methodol*. 2021;21(1):95. doi:10.1186/s12874-021-01285-y
6. Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of acute cardiovascular outcome diagnoses recorded in European electronic health records: a systematic review. *Clin Epidemiol*. 2020;12:1095–1111. doi:10.2147/clep.S265619
7. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One*. 2014;9(3):e92286. doi:10.1371/journal.pone.0092286
8. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One*. 2015;10(8):e0135834. doi:10.1371/journal.pone.0135834
9. Singh S, Fouayzi H, Anzuoni K, et al. Diagnostic algorithms for cardiovascular death in administrative claims databases: a systematic review. *Drug Saf*. 2019;42(4):515–527. doi:10.1007/s40264-018-0754-z
10. Ohlmeier C, Langner I, Hillebrand K, et al. Mortality in the German Pharmacoepidemiological Research Database (GePaRD) compared to national data in Germany: results from a validation study. *BMC Public Health*. 2015;15:570. doi:10.1186/s12889-015-1943-7
11. Ibanez B, James S, Agewall S, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2018;39(2):119–177. doi:10.1093/eurheartj/ehx393
12. (DEGAM) DGfAuFeV. Schlaganfall S3-Leitlinie; 2020. Available from: https://www.awmf.org/leitlinien/detail/ll/053-011.html. Accessed October 21, 2022.
13. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2016;37(27):2129–2200. doi:10.1093/eurheartj/ehw128
14. Deneke T, Borggrefe M, Hindricks G, et al. Kommentar zu den ESC-Leitlinien 2015 "Ventrikuläre Arrhythmien und Prävention des plötzlichen Herztodes". *Der Kardiologe*. 2017;11(1):27–43. doi:10.1007/s12181-016-0115-z
15. Priori SG, Blomstrom-Lundqvist C, Mazzanti A, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: the Task Force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC). Endorsed by: association for European Paediatric and Congenital Cardiology (AEPC). *Eur Heart J*. 2015;36(41):2793–2867. doi:10.1093/eurheartj/ehv316
16. Lix LM, Sobhan S, St-Jean A, et al. Validity of an algorithm to identify cardiovascular deaths from administrative health records: a multi-database population-based cohort study. *BMC Health Serv Res*. 2021;21(1):758. doi:10.1186/s12913-021-06762-0