VALIDATION STUDY PROTOCOL (OP0007)

FOR THE

EUROPEAN NON-INTERVENTIONAL POST-AUTHORIZATION SAFETY STUDY RELATED TO SERIOUS CARDIOVASCULAR EVENTS OF MYOCARDIAL INFARCTION AND STROKE AND ALL-CAUSE MORTALITY FOR ROMOSOZUMAB BY THE EU-ADR ALLIANCE (OP0004)

AND

EUROPEAN NON-INTERVENTIONAL POST-AUTHORIZATION SAFETY STUDY RELATED TO SERIOUS INFECTIONS FOR ROMOSOZUMAB BY THE EU-ADR ALLIANCE (OP0006)

Final Validation Study Protocol	17 Jun 2020

STUDY INFORMATION

Title	Validation study protocol for the "European non-interventional post-authorization safety study related to serious cardiovascular events of myocardial infarction and stroke and all-cause mortality for romosozumab by the EU-ADR Alliance (OP0004)" and the "European non-interventional post-authorization safety study related to serious infections for romosozumab by the EU-ADR Alliance (OP0006)"
Protocol version identifier	1.0
Date of last version of protocol	Not applicable
European (EU) Post-Authorization Study (PAS) register number	To be completed after the protocol is approved
Active substance	Romosozumab
Medicinal product	Romosozumab (AMG785; CDP7851)
Product reference	H0004465
Marketing authorization holder(s)	UCB Pharma S.A.
Joint PASS	Not applicable
Research question and Objectives	Validation of the outcomes of cardiovascular death, myocardial infarction, stroke, and serious infections in patients eligible for inclusion in the 2 studies (OP0004 and OP0006) in different European databases
Countries of study	Denmark, France, Germany, Italy, Netherlands, Spain and the UK
Authors	[redacted]

MARKETING AUTHORIZATION HOLDER

Marketing authorization holder (MAH)	UCB Pharma S.A. Allée de la Recherche 60 B-1070 Brussels Belgium
MAH contact person	[redacted]

1 TABLE OF CONTENTS

LIST OF TABLES

2 LIST OF ABBREVIATIONS

Abbreviation	Definition
ALN	alendronate or alendronic acid
ATC	Anatomical Therapeutic Chemical
CI	confidence interval
CPRD	Clinical Practice Research Datalink
CV	cardiovascular
EMR	electronic medical record
GePaRD	German Pharmacoepidemiological Research Database
GP	general practitioner
HSD	Health Search Database
ICD-10	International Classification of Diseases, revision 10
ICD-10-GM	International Classification of Diseases, revision 10, German modification
ICD-9	International Classification of Diseases, revision 9
ICD-9-CM	International Classification of Diseases, revision 9, clinical modification
ICPC	International Classification of Primary Care
IPCI	Integrated Primary Care Information Project
MI	myocardial infarction
PAS	Post-Authorization Study
PASS	post-authorization safety study(ies)
PPV	positive predictive value
SI	serious infection
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
SNDS	Système National Des Données de Santé
WHO	World Health Organization

3 RESPONSIBLE PARTIES

Function	Name	Title	Affiliation	Address
Principal Investigator	[redacted]	[redacted]	CSM-NDORMS, University of Oxford	Wellington Square, Oxford, OX1 UK
EU and Global Regulatory Lead - Contact Person (MAH)		BSc, LLB, LLM	UCB Biopharma SRL	Allée de la Recherche 60 B-1070 Brussels Belgium
Real World Evidence Lead Scientist - Contact Person (MAH)		Dr (MD, PhD)	UCB Biopharma SRL	Allée de la Recherche 60 B-1070 Brussels Belgium
Safety Lead Romosozumab - Contact Person (MAH)		MSc	UCB Biopharma SRL	Allée de la Recherche 60 B-1070 Brussels Belgium
Databases and local investigators				
CPRD	[redacted]	[redacted]	CSM-NDORMS, University of Oxford	Wellington Square, Oxford, OX1, UK
HSD		[redacted]	SIMG, Società Italiana di Simg Service SRL	Via del Sansovino 179 50142, Florence, Italy
SIDIAP		Dr. (MD, PhD)	IDIAP Jordi Gol	Gran Via de les Corts Catalanes, 587, 08007 àtic, Barcelona, Spain
IPCI		Ass. Prof.	Erasmus University Medical Centre	Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
Nationwide linked Danish registries		Prof.	Department of Clinical Epidemiology at Aarhus University Hospital	Olof Palmes Alle 43-45 8200 Aarhus N Denmark
GePaRD		Dr.	Leibniz-Institut für Präventionsforschung und Epidemiologie (BIPS)	Achterstraße 30 28359 Bremen Germany

Function	Name	Title	Affiliation	Address
SNDS	[redacted]	Dr.	Bordeaux PharmacoEpi Bordeaux (Université de Bordeaux-Adera)	146 rue Léo Saignat - Case 41 33076 Bordeaux cedex, France

CPRD=Clinical Practice Research Datalink; CSM=Centre for Statistics in Medicine; GePaRD=German Pharmacoepidemiological Research Database; HSD=Health Search Database; IPCI=Integrated Primary Care Information Project; MAH=Marketing Authorization Holder; NDORMS=Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences; SIDIAP=Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària; SIMG=Societá Italiana di Medicina Generale; SNDS=Système National Des Données de Santé

4 ABSTRACT

Title

Validation study protocol for the "European non-interventional post-authorization safety study related to serious cardiovascular events of myocardial infarction and stroke and all-cause mortality for romosozumab by the EU-ADR Alliance (OP0004)" and the "European non-interventional post-authorization safety study related to serious infections for romosozumab by the EU-ADR Alliance (OP0006)."

To validate the outcomes of cardiovascular (CV) death, myocardial infarction (MI), stroke, and serious infections (SIs) in patients eligible for inclusion in the 2 post-authorization safety studies (PASS) (OP0004 and OP0006) in different European databases.

Rationale and background

Correct ascertainment of outcomes, exposures, and confounders is essential in pharmacoepidemiological studies. This is particularly important in studies based on routinely collected healthcare data, where diagnosis codes or drug codes are used to identify events of interest. Code validation is performed to measure potential misclassification, allowing for the subsequent exclusion of non-confirmed cases and to ensure the validity of the study outcomes.

Research question and objectives

The objective of this study is to determine the completeness and diagnostic validity of the following outcomes, which will be evaluated in the 2 European non-interventional PASS of romosozumab (OP0004 and OP0006).

For the European Union Post-Authorization Study (EU PAS) related to serious CV events in romosozumab patients (CV PASS), the following outcomes will be validated:

1. CV death
2. MI
3. Stroke

All-cause mortality, another outcome to be evaluated in the CV PASS, has either been previously validated or information is available through database linkage for all databases.

For the EU PAS related to SIs in romosozumab patients (SI PASS), the outcome of

4. SI

will be validated.

Outcomes of interest

Outcomes will be identified based on predefined code lists, which were adapted according to the relevant coding system for the respective database.

Validation algorithms are defined individually for each specific outcome. In addition, the validation procedures, eg, manual free text evaluation, will depend on the specific characteristics of the individual databases and are described separately for primary care databases and claims databases.

Validation of the outcomes of interest will be conducted for databases where previous validation demonstrated insufficient positive predictive values (PPVs) <75% (all-case validation), or where validation has not yet been performed (sample validation).

Population

For the primary care databases, patients eligible for inclusion in the 2 PASS protocols (users of either romosozumab [study drug] or alendronate [ALN; active comparator]) will be identified.

For the German Pharmacoepidemiological Research Database (GePaRD) (Germany), validation will be performed in the general population of the database, which will include a subset of women eligible for inclusion in the 2 PASS.

Data sources

Validation of the outcomes of interest will be carried out in 4 of the 7 longitudinal European healthcare databases, namely Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) (Spain), Health Search Database (HSD) (Italy), Integrated Primary Care Information Project (IPCI) (Netherlands) and GePaRD (Germany). No validation will be performed for the Clinical Practice Research Datalink (CPRD) (UK), Système National Des Données de Santé (SNDS) (France) and the National linked Danish Registers (Denmark).

Data analysis

The PPVs for each outcome will be calculated for each database.

Study size

Sample validation: To estimate an expected PPV of 0.80 with a 95% confidence and a precision of ±0.05, a random sample of 250 cases is required for each outcome.

For all-case validation, all potential cases will be validated.

Milestones

For sample validation, a final report will be submitted after completion of the sample validation studies in year 1 of the study period. Reports for all-case validation will be submitted annually for 6 years, followed by the final report at the end of this period.

5 AMENDMENTS AND UPDATES

Protocol changes may affect the legal and ethical status of the study and may also affect the statistical evaluations of sample size and the likelihood of the study fulfilling its primary objective.

Significant changes to the protocol will only be made as an amendment to the protocol and must be approved by UCB, institutional review board/independent ethics committee, regulatory authorities and local institutions (if required), prior to being implemented.

6 MILESTONES

The validation studies will be implemented in parallel to the CV PASS and SI PASS.

Milestones	**Planned dates***
Registration in the EU PAS register	DD/MMM/YYYY
Start of data collection	
Sample validation (year 1)	
Final sample validation report	
Identification of required all-case validation	
All-case validation (year 1-6)	
Annual interim reports	
Final all-case validation report	

EU=European Union; PAS=Post-Authorization Study
* The dates for the milestones will be included after the protocol is approved.

7 BACKGROUND

Correct ascertainment of outcomes, exposures, and confounders is essential in pharmacoepidemiological studies. This is particularly important in studies based on routinely collected healthcare data, where diagnosis codes or drug codes are used to identify events of interest. Code validation is performed to reduce potential misclassification.

The CV PASS and SI PASS will be conducted in 7 different European healthcare databases, of which 5 are participating in the EU-ADR Alliance:

- CPRD (UK)
- GePaRD (Germany)
- HSD (Italy)
- IPCI (Netherlands),

- National linked Danish Registries, Denmark
- SIDIAP (Spain)
- SNDS (France)

For each database, it was assessed whether the outcomes of interest have already been validated in previous studies. Aside from published reports, results from validation studies in the context of previous PASS were considered.

Some of the outcomes of interest have already been validated in previous studies: For databases, in which previous validation studies demonstrated sufficient completeness and diagnostic validity, eg, high PPVs, no further validation will be required. For CPRD (UK), the National linked Danish Registries (Denmark) and SNDS (France), all outcomes were either previously validated with results demonstrating adequate identification for the respective outcomes or no validation was required (eg, as linked databases, such as mortality records, are available). Thus, no further validations need to be performed in those databases for this validation study.

All-cause mortality has either been previously validated or information is available through database linkage for all databases.

As validation of all events is logistically challenging, sample validation of 250 cases per event will be used for outcomes where no previous validation has been performed. For outcomes, in which previous validation demonstrated insufficient PPV <75% (point estimates), all cases need to be validated.

Table 1 displays the PPVs (%) retrieved from previous studies.

Table 1: Summary of PPVs from previous studies

	MI	Stroke	All-cause mortality	CV death	SI
SIDIAP (ES)	91.3% [a]	75.7% [a]	>90%	53.1% [b]	
CPRD (UK)	92.2% [c]	>89% [d]	NA	NA	Validated [e, f, g]
SNDS (FR)	85.0% [h]	>88% [i]	NA	NA	97% [j]
HSD (IT)	96.6% ('best case') [k] 59.9% ('worst case') [k]		100% [a]	46% [b]	
IPCI (NL)	75.0% ('best case') [k] 46.5% ('worst case') [k]	89.7% [a]	NAP	38% [b]	
Nationwide linked Danish Registries (DK)	>93% [k, l]	>80% [m]	NA	NA	98% [n]
GePaRD (DE)			>83% [o, p]		

CPRD=Clinical Practice Research Datalink; CV=cardiovascular; DE=Germany; DK=Denmark; ES=Spain; FR=France; GePaRD=German Pharmacoepidemiological Research Database; GP=general practitioner; HSD=Health Search Database; IPCI=Integrated Primary Care Information Project; IT=Italy; MI=myocardial infarction; NIS=non-interventional study; NL=Netherlands; PASS=post-authorization safety study(ies); PPV=positive predictive value; SI=serious infection; SIDIAP=Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària; SNDS=Système National Des Donnes de Santé

Note: Colour code: Pink=All-case validation to be performed, Gray=Sample validation to be performed.

Note: Best case scenario=non-assessable cases and non-retrievable cases are not included in the numerator or denominator when the PPV is calculated. Worst case scenario=both non-assessable cases and non-retrievable cases are included in the denominator when the PPV is calculated.

Note: NA=Validation not necessary as information obtained directly from mortality records.

Note: NAP: information on death in IPCI is provided by the GP as 1 of the reasons why a patient left the practice. In addition, by means of Natural Language Processing, additional searches for mortality are done each time a new dataset is released to pick up additional deaths. Specificity is 100% - Sensitivity might be lower but there is no gold standard to validate missing deaths in the database.

[a] EUPAS7674 CA. NIS Final Report, 2019; [b] Available from the data owner based on other PASS-related activity in partnership with EU-ADR Alliance; [c] Herrett et al, 2013; [d] Andersohn et al, 2006; [e] Aberra et al, 2007; [f] Jick et al, 2006; [g] van Staa and Abenhaim, 1994; [h] Bezin et al, 2015; [i] Giroud et al, 2015; [j] Sahli et al, 2016; [k] Coloma et al, 2013; [l] Sundbøll et al, 2016; [m] Krarup et al, 2007; [n] Holland-Bill et al, 2014; [o] Ohlmeier et al, 2015; [p] Ohlmeier et al, 2016

CV death

Cardiovascular death has been validated in a previous EU-ADR PAS validation study in SIDIAP (Spain), HSD (Italy) and IPCI (Netherlands), with PPVs ranging between 38% and 53.1%. For CPRD (UK), Nationwide linked Danish Registries (Denmark) and SNDS (France), the cause of death can be obtained directly from mortality records, which is considered to be the gold standard.

MI

A very high validity of the coding of MI has been shown in the literature: PPVs ranging from 75% to 100% within EU-ADR Alliance partners were shown (Coloma et al, 2013). The PPVs varied slightly depending on the coding system of the diagnoses used. Specific lists were

developed and evaluated in previous validation studies, performing as follows: International Classification of Diseases, revision 10 (ICD-10) codes had a PPV of 100%; International Classification of Diseases, revision 9, clinical modification (ICD-9-CM) codes had a PPV of 96.6% (95% confidence interval [CI]: 93.2%-99.9%) and International Classification of Primary Care (ICPC) codes had a PPV of 75% (95% CI: 67.4%-82.6%) (Coloma et al, 2013). Database-specific validation studies have also been conducted, suggesting similar findings (Madsen et al, 2003). Besides published estimates of the PPVs, PPV assessments performed in the context of other PASS by the data owners were also considered. For SNDS (France), a validation study has shown PPVs ranging from 67% to 100% depending on the chosen ICD-10 codes (Bezin et al, 2015).

Stroke

Stroke diagnoses have been validated in Denmark, with a PPV ranging from 80.5% to 86% (Krarup et al, 2007). For SIDIAP (Spain) and IPCI (Netherlands), 'stroke' has been validated in a previous EU-ADR study, demonstrating PPVs of 75.7% and 89.7%, respectively. Giroud and colleagues (2015) validated French hospital discharge codes for stroke and transient ischemic attack retrieving PPVs around 90% for the different algorithms they developed. In a CPRD-based study (former GPRD), 90% of computerized diagnoses of stroke could be confirmed after reviewing written records of a sample of patients (Andersohn et al 2006, citing Hall et al, 2006). While no extensive validation has been performed yet, stroke has been investigated using "external rate" comparison methods in GePaRD (Germany) (Schink et al, 2018).

SIs

Code validation for SIs for all databases has been reported previously for CPRD (UK) and the Nationwide linked Danish Registers (Denmark): Holland-Bill and colleagues conducted a validation study using the Nationwide linked Danish Registers (DK) in 2014. For CPRD (UK), a number of validation studies have been reported on different types of specific infections including tuberculosis (Aberra et al, 2007; Jick et al, 2006), impetigo, measles, rubella, scabies and viral illness (De Wilde et al, 2004) as well as urinary tract infection (De Wilde et al, 2004; van Staa and Abenhaim, 1994), and respiratory infection (van Staa and Abenhaim, 1994). For SNDS, infections were validated using hospital discharge diagnoses in the French national hospital database (Sahli et al, 2016).

This operational protocol describes the proposed validation algorithms for each outcome in detail and outlines the process of transformation of the input files to the validation dataset.

8 RESEARCH QUESTION AND OBJECTIVES

The objective of this study is to determine the completeness and diagnostic validity of the following outcomes, which will be evaluated in the 2 European non-interventional PASS of romosozumab.

For the EU PAS related to serious CV events in romosozumab patients (CV PASS), the following outcomes will be validated:

1. CV death
2. MI
3. Stroke

All-cause mortality, another outcome to be evaluated in the CV PASS, has either been previously validated or information is available through database linkage for all databases.

For the EU PAS related to SIs in romosozumab patients (SI PASS), the outcome of

4. SI

will be validated.

9 RESEARCH METHODS

9.1 Outcome definitions

Validation studies will be conducted for each of the following study outcomes:

1. CV death
2. MI
3. Stroke
4. SI

CV death

For this study, CV death is defined as death in the context of an event of 'sudden cardiac death', 'heart failure', cardiac arrhythmia', 'stroke'/'cerebrovascular disease', or 'myocardial infarction'/'heart attack'.

MI

Myocardial infarction is one of the manifestations of coronary heart disease. It is defined by the demonstration of myocardial cell necrosis due to significant and sustained restriction in blood supply that causes a lack of oxygen delivery to myocardial tissue (ischemia). It often presents as an acute manifestation of atherosclerosis-related coronary heart disease. The clinical definition denotes the presence of acute myocardial injury detected by abnormal cardiac biomarkers, typically cardiac troponin, in the setting of evidence of acute myocardial ischemia. Myocardial infarction may be classified based on electrocardiogram results, ST-elevation myocardial infarction (STEMI) or Non-ST elevation myocardial infarction (NSTEMI) (Thygesen et al, 2018; Mendis et al, 2010).

Stroke

A 'stroke' is typically characterized as a neurological deficit, which is attributed to an acute focal injury of the central nervous system by a vascular cause including cerebral infarction and both

intracerebral and subarachnoid haemorrhage (Sacco et al, 2013). The World Health Organization (WHO) defines stroke as "rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin *(Sacco et al., Stroke 2013)*." For the CV PASS, a broad definition of stroke including both ischemic and non-ischemic strokes will be applied. Transient ischemic attacks, with symptoms resolving within 24 hours, will be excluded.

SIs

Serious infections are defined for the SI PASS as recording of 1 or more predefined infection code(s) with evidence of hospital admission (primary outcome) or related death (secondary outcome).

For the primary outcome:

1. Presence of 1 or more infection codes recorded as a diagnosis in a hospital admission where hospital or claims data are available.
2. One or more infection codes identified from primary care records with evidence of a related hospital admission (identified from free text).

For the secondary outcome:

3. In-hospital death with a diagnosis of infection during the same admission.
4. Recorded diagnosis of any infection in either primary or secondary care/hospital records, followed by death in the subsequent month.

9.2 Validation methods

Validation algorithms are defined individually for each specific outcome. In addition, the validation procedure, eg, manual free text evaluation, will depend on the specific characteristics of the individual databases. In the following sections, validation procedures are described separately for primary care databases (SIDIAP [Spain], HSD [Italy], and IPCI [Netherlands]) and claims databases (GePaRD [Germany]).

9.2.1 Validation in primary care databases

9.2.1.1 Case identification

Patients eligible for inclusion in the comparative safety studies (users of either romosozumab [study drug] or ALN [active comparator]) will be identified.

Outcomes will be identified based on predefined code lists, which were adapted according to the relevant coding system for the respective databases. A detailed code list for the identification of MI and stroke are provided in Appendix 3 of the CV PASS protocol and in Appendix 2 (Table A) of this protocol. Preliminary code lists for the identification of "infection" are also provided in Appendix 2 (Table B) of this protocol.

For CV death, death will be identified from the primary care records or linked mortality registers. Depending on the individual database, CV events that were recorded around the time of death will be identified using predefined code lists or keywords for free text review (Appendix 2, Table C).

9.2.1.2 Sample validation

For databases where the respective outcome was not previously validated, sample validation will be performed among patients eligible for inclusion in the CV PASS and SI PASS protocols. Sample validation has been conducted in previous EU-ADR validation studies as validation of all events is logistically challenging. This is a very efficient method to evaluate the diagnostic validity of the coded outcomes in the study population, as results obtained from a random sample are representative for the cohort from which the sample was drawn.

For each database, a random sample of 250 eligible cases will be taken for each outcome of interest. Samples will be retrieved from the study cohort of users of romosozumab (study drug) or ALN (active comparator) over the course of the first year of the study periods of the CV PASS and SI PASS, respectively.

Random sampling of cases will be carried out automatically by the Jerboa data analysis tool of the EU-ADR Alliance as illustrated below. This tool uses common input files, extracted by each data custodian locally, and produces a dataset containing the events to be validated.

DB=database

A random sample of 246 cases is needed to estimate an expected PPV of 0.80 with a 95% confidence and a precision of ±0.05. In case less than 250 cases of an outcome will be identified in the first year, data from subsequent years will be included to sample cases.

9.2.1.3 All-case validation

All-case validation will be conducted when outcomes were previously validated in the respective database, but an insufficient PPV of <75% (point estimate) was demonstrated (Table 1).

For these outcomes/databases, all patients who i) are eligible for inclusion in the CV PASS/ SI PASS comparative safety analysis and ii) experienced the respective outcome, will be extracted. Each "potential" case will then be validated individually by trained medical reviewers who are blinded to exposure. Only those "potential" cases that are confirmed as cases during validation will be included in the comparative safety analyses. For databases where reviewers cannot be blinded, alternative measures to reduce potential bias will be taken.

9.2.1.4 Validation of CV death

While HSD, IPCI, and SIDIAP contain information on date of death, cause of death is not automatically recorded.

All eligible cases of CV death in the study cohorts in SIDIAP (Spain), HSD (Italy) and IPCI (Netherlands) will need individual validation before they are included in the comparative safety analyses.

Cardiovascular death will be identified by a combination of the predefined codes/terms for 'sudden cardiac death', 'heart failure', 'cardiac arrhythmia', 'stroke'/'cerebrovascular disease' or 'myocardial infarction'/'heart attack', followed by death, with date of death recorded within ±60 days of the latest code for any of the aforementioned events.

All validations will be based on free text review of the individual cases. Primary care charts and any related documents (eg, specialist letters, referrals) will be reviewed by clinically trained validators blinded to exposure. The algorithm for the validation of CV death is provided in Appendix 3. Adaptations and specifications to the algorithm may be required to adapt the algorithm to database-specific needs.

9.2.1.5 Validation of MI

All eligible cases of MI in IPCI (Netherlands) and HSD (Italy) will need individual validation.

All validations will be based on free text review of the individual cases in a time window of 30 days before/after the event date. Primary care charts and any related documents (eg, specialist letters, referrals) will be reviewed by clinically trained validators using prespecified algorithms and blinded to exposure. The algorithm has been applied in a previous EU-ADR validation study and is provided in Appendix 4. Adaptations and specifications to the algorithm may be required to adapt the algorithm to database-specific needs.

9.2.1.6 Validation of stroke

Sample validation of 250 stroke cases in HSD (Italy) will be required.

Sample validation in HSD (Italy) will be based on free text review of the individual cases in a time window of 30 days before/after the event date. Primary care charts and any related documents (eg, specialist letters, referrals) will be reviewed by clinically trained validators using a prespecified algorithm (provided in Appendix 5), which has been adapted from an algorithm used in a previous EU-ADR validation study. Additional adaptations and specifications to the algorithm may be required to adapt the algorithm to HSD (Italy) needs.

9.2.1.7 Validation of SI

Validation of the diagnosis of infection/s will be conducted in SIDIAP (Spain), HSD (Italy) and IPCI (Netherlands).

Serious infections are defined for the SI PASS as recording of 1 or more infection code(s) with evidence of hospital admission or related death:

1. Presence of 1 or more infection codes recorded as a diagnosis in a hospital admission where hospital or claims data are available.
2. One or more infection codes identified from primary care records with evidence of a related hospital admission (identified from free text).
3. In-hospital death with a diagnosis of infection during the same admission.
4. Recorded diagnosis of any infection in either primary or secondary care/hospital records, followed by death in the subsequent month.

In the SI PASS, SI with evidence of hospital admission is defined as the primary objective, whereas SI followed by death is the secondary objective.

As the definition of SI includes information on hospitalization and death, the validation requirements and relevant definitions will differ between primary care databases with respect to the availability of linked hospital and mortality data.

Recording of mortality

All-cause death has been previously validated in SIDIAP (Spain), HSD (Italy), and IPCI (Netherlands).

Databases with hospital/claims data linkage (SIDIAP)

Hospital admissions will be identified through linkage to the loco-regional hospital admissions data in the Mínimo de Datos (CMBD) database.

As SI has not been previously validated in SIDIAP (Spain), sample validation studies of 250 cases relevant for the primary objective will be needed for the diagnosis of infection/s. In case point estimates for PPV <75% will be obtained from sample validations, subsequent all-case validation needs to be conducted.

Relevant definition for primary objective:

1. Presence of 1 or more infection codes recorded as a diagnosis in a hospital admission where hospital or claims data are available.

Relevant definition for secondary objective:

4. Recorded diagnosis of any infection in either primary or secondary care/hospital records, followed by death in the subsequent month.

Databases without hospital/claims data linkage (HSD, IPCI)

Although SI has not been previously validated in HSD (Italy) and IPCI (Netherlands), validation of only a sample of the patients will not be possible due to the lack of availability of linkage to either hospital or claims data.

The HSD (Italy) and IPCI (Netherlands) will review free text for all potential 'infection' cases for the comparative safety analysis (amongst eligible users of romosozumab or ALN in the study period) to identify related hospital admissions for the identification of potential cases.

Medical charts, hospital letters and any other documents available will be reviewed by medical students and/or medical doctor blinded to exposure.

Relevant definition for primary objective:

2. One or more infection codes identified from primary care records with evidence of a related hospital admission (identified from free text).

Relevant definition for secondary objective:

4. Recorded diagnosis of any infection in either primary or secondary care/hospital records, followed by death in the subsequent month.

Recording of infection

The list of ICD-10 codes from a previous Danish validation study by Holland Bill et al (2014) will be used to identify diagnoses for infection.

These codes will be mapped equivalent to the International Classification of Diseases, revision 9 (ICD-9)/ICD-9-CM- and ICPC-coding systems, allowing for the use of the same definitions in SIDIAP (Spain), HSD (Italy) and IPCI (Netherlands). The preliminary code list is included in Appendix 2 (Table B) of this protocol; additional codes may be added, if necessary.

As mentioned above, different definitions of SI will be used owing to the individual characteristics of the different databases. The applied SI definition will be recorded for each case, to enable calculation of PPV separately for SI related to hospitalization and SI related to death. No individual validation of specific infection subtypes (eg, respiratory infection) will be performed.

Procedure

Potential cases of 'infection' will be identified based on predefined code lists. Related hospitalizations or death will be identified to characterize the infections as "serious".

Validation of SI in primary care databases (SIDIAP [Spain], HSD [Italy], and IPCI [Netherlands]) will be based on free text review of the individual cases in a time window of 30 days before/after the event date.

Primary care charts and any related documents (eg, specialist letters, referrals) will be reviewed by clinically trained validators – blinded to exposure – using prespecified chart extraction forms and assessment sheets as, for example, used by Holland-Bill et al (2014) for their validation study in Denmark. The forms will be adapted to the specific needs for the individual databases in collaboration with medical experts where necessary.

9.2.2 Validation in claims databases

9.2.2.1 Case identification

Sample validation in the GePaRD (Germany) database will be performed retrospectively in the general database. A subcohort of women who are eligible for study inclusion (eg, age) will be included in the overall sample, allowing conclusions to be drawn for this group. For this, a random sample may be taken locally without using the Jerboa tool.

Outcomes will be identified based on predefined code lists for MI, stroke, and SI, which are provided in Appendix 2 (Table A and Table B) of this protocol.

Cardiovascular death will be identified by a combination of the predefined codes for 'sudden cardiac death', 'heart failure', 'cardiac arrhythmia', 'stroke'/'cerebrovascular disease', or 'myocardial infarction'/'heart attack', followed by death, with date of death recorded around the time of any of the aforementioned events. The International Classification of Diseases, revision 10, German modification (ICD-10-GM) code lists for the identification of the CV events of interest are provided in Appendix 2 (Table C).

9.2.2.2 Sample validation and all-case validation

As none of the outcomes of interest, namely CV death, MI, stroke, and SI, have been previously validated in GePaRD, sample validation studies of 250 cases will be needed to validate each of the respective outcomes.

All-case validation will be conducted for all outcomes where insufficient PPV of <75% (point estimates) is demonstrated in the sample validation.

In that case, all patients who i) are eligible for inclusion in the CV PASS/SI PASS comparative safety analysis and ii) experienced the respective outcome, will be extracted. Each “potential” case will then be validated individually by trained medical reviewers who are blinded to exposure. Only those "potential" cases that are confirmed as cases during validation will be included in the comparative safety analyses.

9.2.2.3 Validation of CV death, MI and stroke

The validation of 250 sample cases in GePaRD (Germany) will be performed based on a review of samples of individual case profiles, that is, all information (eg, diagnoses, dispensations, procedures, hospitalized time) of the cases around the time of the event will be reviewed independently by 2 medical experts. They will independently adjudicate the case profiles using prespecified algorithms. Inter-rater reliability will be described. Discordant assessments will be resolved by consensus or a third expert and seen as an area for further training. Validation algorithms used in the primary care databases (Appendices 3-5) will be adapted as closely as possible to a secondary data case profile algorithm in collaboration with medical experts.

9.2.2.3.1 Validation of SI

Serious infections are defined for the SI PASS as recording of 1 or more infection code(s) with evidence of hospital admission or related death. The ICD-10 codes used by Holland-Bill et al (2014) (Appendix 2, Table B) were mapped equivalent to the ICD-10-GM, allowing the use of the same definitions as for the primary care databases. Dates for hospitalization admission as well as discharge diagnoses and mortality can be accessed directly from the database. Hospital discharge diagnoses will be used to identify SI.

Relevant definition for primary objective:

1. Presence of 1 or more infection codes recorded as a diagnosis in a hospitalization where hospital or claims data are available.

Infection codes recorded as a hospital discharge diagnosis will be identified and cases, in which the infection started during the hospitalization, will be excluded where feasible.

Relevant definition for secondary objective:

3. In-hospital death with a (discharge) diagnosis of infection during the same hospitalization.

For sample validation, a total of 250 cases relevant for the primary objective will be extracted. The applied SI definition will be recorded for each case, to enable calculation of PPV separately for SI related to hospitalization and SI related to death.

Stratification will not be performed for specific infection subtypes (eg, respiratory infection).

Medical chart extracti,on forms and assessment sheets used in the primary care databases (Holland-Bill et al, 2014) will be adapted as closely as possible to a secondary data case profile algorithm in collaboration with medical experts.

9.3 Data sources

Validation of the outcomes presented in Section 9.1 will be carried out in 4 of the 7 longitudinal European healthcare databases (excluding CPRD [UK], SNDS [FR] and the National linked Danish Registries [DK]). The 7 European healthcare databases are listed in Section 7 and brief descriptions of GePaRD (Germany), HSD (Italy), IPCI (Netherlands) and SIDIAP (Spain) databases are provided in this section.

As described earlier in Section 9.2.1 and Section 9.2.2, different approaches to the validation process will be applied, depending on the characteristics of the database including availability of linkage to data from secondary or tertiary care and death registries.

Table 2 provides an overview of the characteristics of the 7 individual databases, each of which are described in detail in the CV PASS and SI PASS protocols.

Table 2: Database characteristics

Database	CPRD	GePaRD	HSD	IPCI	National linked Danish Registries	SIDIAP	SNDS
Country	UK	DE	IT	NL	DK	ES	FR
Type of database	MR	ADM	MR	MR	ADM	MR	ADM
	Primary care	Health insurance claims	Primary care	Primary care	National registries	Primary care	Outpatient claims, hospital discharge summaries, death registries
No. active patients in millions.	5	25	1.5	1.2	5.8	5.8	66
Outpatient diagnoses	Yes	Yes	Yes	Yes	Yes (hospital outpatient)	Yes	Yes
Coding system for diagnoses	READ	ICD-10-GM	ICD-9-CM	ICPC	ICD-10	ICD-10 (ICD-9 for hospital diagnoses)	ICD-10
Hospitalization	Yes (60%, linked to HES)	Yes	Yes	Yes	Yes	Yes (30%)	Yes
Date of death	Yes	Yes	Yes	Yes	Yes	Yes	Yes (mm/yyyy)
Cause of death	Yes, from ONS linkage (60%)	No	No	No (available from free text)	Yes (lag 2 years)	No	Yes (lag 2- to 4-years)

ADM=administrative; CM=clinical modification; CPRD=Clinical Practice Research Datalink; DE=Germany; DK=Denmark; ES=Spain; FR=France; GePaRD=German Pharmacoepidemiological Research Database; GM=German modification; HES=Health Episode Statistics; HSD=Health Search Database; ICD=International Classification of Diseases; ICD-10=ICD, revision 10; ICD-9=ICD, revision 9; ICPC=International Classification of Primary Care; IPCI=Integrated Primary Care Information Project; IT=Italy; MR=medical records; NL=Netherlands; ONS=Office for National Statistics; SIDIAP=Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària; SNDS=Système National Des Données de Sante

The following descriptions of the 4 databases, in which validation studies will be performed, are taken from the CV PASS and SI PASS protocols:

GePaRD - GERMANY

The GePaRD consists of claims data from 4 German statutory health insurance providers covering approximately 25 million individuals throughout Germany. Cross-sectionally, GePaRD covers about 15 million individuals, which represents approximately 17% of the German population of 82 million inhabitants.

The GePaRD contains individual level information on demographic characteristics, hospitalizations (including admission diagnoses, main discharge diagnoses, and reason for discharge including death), outpatient physician visits, and outpatient drug dispensations for reimbursed products. While exact dates are provided for hospitalizations and outpatient drug dispensations, only the year and quarter are known for outpatient diagnoses. Drugs that are purchased over the counter are not contained in the database. With a few exceptions, the same applies to medication administered in hospital. The acceptability of GePaRD for pharmacoepidemiological research has been assessed methodologically as well as by validation studies (Ohlmeier et al, 2016; Ohlmeier et al, 2015; Ohlmeier et al, 2014; Pigeot and Ahrens, 2008). Recently, GePaRD has been used for various types of pharmacoepidemiological studies including drug utilization studies and studies investigating the risks of drugs or vaccines (Schink et al, 2018; Schmedt et al, 2016a; Schmedt et al, 2016b; Schink et al, 2014).

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 (Social) Insurance Office and the Senator for Science, Health, and Consumer Protection in Bremen as the responsible authorities, approved the use of GePaRD data for the PASS. Informed consent for studies based on GePaRD is not required by law, and according to the Ethics Committee of the University of Bremen, these studies are exempt from IRB review.

HSD - ITALY

The Italian partner for the study will use the HSD, a longitudinal observational database that is representative of the Italian general population. The HSD was established in 1998 by the Italian College of General Practitioners (Filippi et al, 2005). The HSD contains data from computer-based patient records from a selected group of general practitioners (GPs) covering a total of 1.5 million patients located throughout Italy. These GPs voluntarily agreed to collect data for the database and attend specified training courses. The database includes information on the age, gender, and identification of the patient and GP registration information, which is linked to prescription information, clinical events and diagnoses, hospital admission, and causes of death. All diagnoses are coded according to ICD-9-CM. Drug names are coded according to the Anatomical Therapeutic Chemical (ATC) classification system (WHO, 2019). To be included in the study, GPs must have provided data for at least 1 year and meet standard quality criteria pertaining to levels of coding, prevalence of well-known diseases, and mortality rates (Cricelli et al, 2003). The HSD has been used as a data source for a number of peer-reviewed publications on the prevalence of disease conditions, drug safety and prescription patterns in Italian primary care (Cazzola et al, 2011). Approval for use of data is obtained from the Italian College of General Practitioners for each study.

IPCI DATABASE-NETHERLANDS

In 1992, the IPCI was started by the Department of Medical Informatics of the Erasmus University Medical Centre. The IPCI is a longitudinal observational database that contains data from computer-based patient records of a selected group of GPs throughout the Netherlands, who voluntarily chose to supply data to the database. Collaborating practices are located throughout the Netherlands and the collaborating GPs are comparable to other GPs in the country according to age and gender. In the Netherlands, all citizens are registered with a GP practice, which forms the point of care and acts as a gatekeeper in a 2-way exchange of information with secondary care. The medical records of each patient can therefore be assumed to contain all relevant medical information including medical findings and diagnosis from secondary care.

The IPCI database is representative of the Dutch population regarding age and gender (Voordouw et al, 2004).

The database contains information on about 2.5 million patients. This is the cumulative number of patients who have ever been part of the dynamic cohort of registered patients. The ICPC is the coding system for patient complaints and diagnoses but diagnoses and complaints can also be entered as free text. Prescription data such as product name, quantity prescribed, dosage regimen, strength and indication are entered into the computer (Vlug et al, 1999). The National Database of Drugs, maintained by the Royal Dutch Association for the Advancement of Pharmacy, enables the coding of prescriptions, according to the ATC classification scheme recommended by WHO (2019). Approval needs to be obtained for each study from the governance board.

SIDIAP DATABASE - SPAIN

The GPs play an essential role in the public healthcare system of Spain as they are responsible for primary health care, long-term prescriptions and specialist and hospital referrals. The Spanish public healthcare system covers more than 98% of the population. The SIDIAP database comprises of electronic medical records (EMRs) of a representative sample of patients attended by GPs in Catalonia (North-East Spain), covering a population of more than 5.8 million patients (over 80% of the total of 7.5 million population of Catalonia) from 274 primary care practices with 3414 participating GPs. The SIDIAP data comprises the clinical and referral events registered by primary care health professionals (GPs and nurses) and administrative staff in EMRs, comprehensive demographic information, prescription and corresponding pharmacy invoicing data, specialist referrals, primary care laboratory test results, and hospital admissions and their major outcomes.

Health professionals gather this information using ICD-10 codes and structured forms designed for the collection of variables relevant for primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, and blood and urine test results. Only GPs who meet quality control standards can participate in the SIDIAP database. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP database. Recent reports have shown the SIDIAP data to be useful for epidemiological research (García-Gil Mdel et al, 2011). Studies performed using SIDIAP data require previous approval by both the Scientific and the Ethics Committee.

9.4 Data management

9.4.1 Input files

For the validation, 3 different types of data files will be extracted and are described in detail in the following sections: Patient files, prescription files and event files. The table shells should serve as examples and may need to be adapted to fit the individual needs of the databases.

9.4.1.1 Patients.txt

This file contains general information on the patients in the database who have at least 1 of the events of interest to validate: CV death, MI, stroke, or SI according to the primary data source.

Sample validation: For each data source (except GePaRD), Jerboa will randomly select 250 patients per event. Ideally, only patients for which the data custodian can provide data for validation (eg, free text, letters or linkage to external data sources) should be included. In case many samples are non-assessable, for example, due to the lack of free text information, the Jerboa tool can oversample to aim for 250 cases per event type.

For all-case validation, all "potential" cases will be extracted.

The patients.txt dataset will contain 1 comma-separated line of data for each patient, containing the following variables:

PatientID	Patient Identifier. *No missing values allowed.*
Birthdate	Date of birth. *No missing values allowed.*
Gender	Gender. Can be either F or M for Female or Male, respectively. *No missing values allowed.*
Startdate	Date from which the patient is eligible to be included in the study. Possible value: yyyymmdd. *No missing values allowed.*
Enddate	Date after which the patient is no longer eligible for inclusion in the study. Possible value: yyyymmdd. *No missing values allowed.*
Link_Hosp	Binary variable on possibility for linkage of hospital records in SIDIAP. Possible values: 0 (no linkage available) or 1 (linkage available).
Link_Mortality	Binary variable on possibility for linkage of hospital records in SIDIAP. Possible values: 0 (no linkage available) or 1 (linkage available).

SIDIAP=Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària

An example for the structure of the dataset "Patients.txt":

PatientID,Birthdate,Gender,Startdate,Enddate,Link_Hosp,Link_Mortality:
100001,19420201, F,20201212,20211230,1,1

9.4.1.2 Prescriptions.txt

Prescriptions.txt contains the following information about prescriptions for the patients included in Patients.txt for either romosozumab (ATC code M05BX06) or ALN (ATC code M05BA04).

PatientID	Patient Identifier. *No missing values allowed.*
ATC	ATC code of the drug (7 characters). *No missing values allowed.*
Date	Date of prescription/dispensation of romosozumab or ALN (dispensation date preferred). Possible values: yyyymmdd. *No missing values allowed.*
Duration	Duration of the prescription in days. The duration of treatment should be calculated by the data custodian using the following definitions: - Total units/prescribed daily dose. - Total number of DDDs prescribed. - Total units*strength/DDD value. - Duration as entered by physicians where applicable. *No missing values allowed.*

ALN=alendronate or alendronic acid; ATC=Anatomical Therapeutic Chemical; DDD=defined daily dose

An example for the structure of the dataset "Prescriptions.txt":
PatientID,ATC,Date,Duration: 100001,M05BX03,20210201,30

9.4.1.3 Events.txt

Events.txt comprises all events of interest, namely CV death, MI, stroke, and SIs for all patients in Patients.txt as illustrated below. These events are defined by the data custodian using the predefined lists of codes.

As death and hospitalization are required for the identification of SI, these events are also included. Multiple lines per patient are possible.

PatientID	Patient Identifier. *No missing values allowed*.	
Date	Date of the event. Possible values: yyyymmdd. *No missing values allowed*.	
EventType	Type of the event. *No missing values allowed*. Possible values:	
	ALLDEATH[a]	All-cause mortality
	CVDEATH	Cardiovascular death
	MI	Myocardial infraction
	STROKE	Stroke
	HOSPITAL[b]	Hospitalization
	(S)I	(Serious) infection
Code	*No missing values allowed.*	

CV=cardiovascular
[a] Date of death to facilitate screening for death on or after infection codes or CV event codes.
[b] Date of hospitalization to facilitate screening for hospitalization on or after infection codes.

An example for the structure of the dataset "Events_CV.txt":
PatientID,Date,Eventtype,Code: 100001,20210201,MI,I21

9.4.2 Validation files

The events for validation will be sampled from the cohorts of users of romosozumab or ALN, who are eligible for inclusion to the comparative safety studies in the CV PASS and SI PASS, respectively.

The first safety incident event during the study period for each event type will be used for validation. Thus, each patient can only contribute 1 event per event type. As for the exclusion criteria for the CV PASS and SI PASS, patients with the history of stroke or MI at any time before therapy initiation are excluded, as well as patients with a history of SI in the 30 days before index date.

Jerboa will create a *Validation.csv* file with the following columns. Multiple lines per patient are possible:

PatientID	Patient Identifier. *No missing values allowed.*	
Date	Date when any of the identified codes for ALLDEATH, CVDEATH, MI, HOSPITALIZATION, STROKE or SI was recorded. Possible values: yyyymmdd. *No missing values allowed.*	
EventType	ALLDEATH, CVDEATH, MI, STROKE, HOSPITALIZATION, or SI possible values:	
	ALLDEATH	All-cause mortality
	CVDEATH	Cardiovascular death
	MI	Myocardial infarction
	STROKE	Stroke
	HOSPITAL	Hospitalization
	SI	Serious infection
EventDate	Empty column to be filled in by the data custodian. Possible values (please use numbers as code): 0=ALTERED 1=CORRECT	
NEW EventDate	Empty column to be filled in by the data custodian. If event dates do not match, please provide new date in yyyymmdd format.	
SI REASON	Empty column to be filled in by the data custodian. Possible values (please use numbers 1-2 as code): 1=Hospitalization for infection (primary objective) 2=Infection followed by death (secondary objective) *If event not SI, please put NA.* *No missing values allowed.*	
Validity	Empty column to be filled in by the data custodian. For each case, the result retrieved using the respective validation algorithm will be noted. Possible values: definite/possible/no/not assessable *No missing values allowed.* *Possible values may differ depending on the event-specific algorithm.*	

NA=not applicable

For GePaRD, random samples will be taken locally from a retrospective patient collective without use of the Jerboa tool. Comparable validation tables will be created.

The validation files can be opened and edited in Excel. The data custodian will fill in the validity columns, the "SI Reason" column (if applicable) and alter the event date (if applicable) manually based on the background information retrieved.

After validation of all cases, the validation files will remain locally with the data custodian.

Output tables will be provided based on the filled validation files and regularly uploaded to the OCTOPUS remote research (where possible with respect to local data protection rules).

9.4.3 Output files

The following tables (Table 3 and Table 4) will be provided for each database after completion of the sample validations and annually for all-cases validations, respectively.

Table 3: Characteristics of the validation process

	CV death	MI	Stroke	SI
N (Total events eligible)				*
N (Events selected for validation)				*

CV=cardiovascular; MI=myocardial infarction; SI=serious infection
* Number of cases labelled serious for primary and secondary objective definition separately.

Table 4: Outcome validation

CV death						
	Sample size	**Definite cases**	**Non-cases**	**Non-assessable cases**	**Event date confirmed**	
CV death						
MI						
	Sample size	**Definite cases**	**Probable cases**	**Non-cases**	**Non-assessable cases**	**Event date confirmed**
MI						
Stroke						
	Sample size	**Definite cases**	**Probable cases**	**Non-cases**	**Non-assessable cases**	**Event date confirmed**
Stroke						
SI						
	Sample size	**Cases**	**Non-cases**	**Non-assessable cases**	**Event date confirmed**	
SI						
1=Hospitalization for infection						
2=Infection followed by death						

CV=cardiovascular, MI=myocardial infarction; SI=serious infection
Note: Please put NA if not applicable.

9.5 Data analysis

Descriptive analyses

Summary descriptive statistics will be provided for validated samples, as well as for all-case validation. A subset of the covariates described in the CV PASS and SI PASS will be selected based on clinical relevance. Covariates will be assessed as defined in the PASS protocols, eg, in the year before/at the index date of the drug-defining exposure status for the patient at the time of the event. In collaboration with the database owners, the best method (eg, linkage) for making the Baseline characteristics, which were extracted during the PASS, available for the validated samples will be discussed.

Baseline characteristics of the validated cases will be assessed and compared to the Baseline characteristics of the overall study.

For GePaRD, summary descriptive statistics comprising general patient characteristics (eg, age, sex, comorbidity, comedication), characteristics relevant for each outcome (eg, risk factors), and characteristics relevant for the study (eg, markers of osteoporosis severity) will be provided for the validated samples, which will be taken from the overall database population.

PPVs

The PPVs for each of the events described above will be calculated for each data source. The PPV is defined as follows:

$$\text{PPV} = \frac{\text{N (confirmed cases)}}{\text{N (confirmed cases)} + \text{N (non} - \text{cases)}}$$

The 95% CIs for PPVs will be estimated assuming a binomial distribution. STATA version 15.1 (StataCorp, College Station, Texas, USA) or R (version 3.2.3, the R Foundation for Statistical Computing) will be used for statistical analysis.

Missing data

Ideally, sample validation will be conducted on samples with available free text information. Oversampling may be considered to retrieve 250 analyzable samples. However, in case of too little information available for assessment, cases will be classified as non-assessable.

9.6 Quality control

This study has been designed and shall be implemented and reported in accordance with the Guidelines for good pharmacoepidemiology practices of the International Society for Pharmacoepidemiology (ISPE, 2015), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (von Elm et al, 2008), and with the ethical principles laid down in the Declaration of Helsinki.

Clinically trained validators, blinded to exposure, will review cases using prespecified algorithms. For claims data, cases will be reviewed independently by 2 medical experts. They will independently adjudicate the case profiles using prespecified algorithms. Discordant assessments will be resolved by consensus or a third expert.

9.7 Limitations of the research methods

This is a study to validate the outcomes of CV death, MI, stroke, and SIs in patients eligible for inclusion in the 2 studies (OP0004 and OP0006) in different European databases or in the general database population, respectively.

Case identification will be based on prespecified codes. Thus, identified "potential" cases will be validated, but no additional cases will be identified during the validation process.

Validation in the primary care databases will be performed based on free text, and thus relies on the recording of information by GPs and the availability of specific information in the database.

9.8 Other aspects

Not applicable

10 PROTECTION OF HUMAN SUBJECTS

This validation study will use secondary data collection and does not pose any risks for patients. Participants from 4 different European member states will process individual data as collected in national electronic health record databases in compliance with all applicable national and European regulation as well as with ethical and regulatory issues including those on privacy.

All of the databases used in this study are already used for pharmacoepidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to.

In agreement with these regulations, local analyses will be run, which generate non-identifiable data with less information. Validation files that include PatientIDs and dates will remain at the local study site and will be stored safely in accordance with the European and local data protection regulations.

The output files will be stored in a central remote research environment held by Erasmus MC, where possible. These output files do not contain any data that allow identification of patients included in the study. In fact, each record is completely anonymous and does not contain any identifier key. Starting from this, the remote research environment implements further security measures in order to ensure a high level of stored data protection, according with the article 34 of legislative decree 196/2003 and article 22 of Regulation (EC) 45/2001.

The study will be registered in the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) EU PAS Register.

In addition, a scientific advisory committee consisting of external experts will be constituted to guarantee scientific soundness of the study and also to follow-up on the progress and the appropriate conduct of the study.

11 MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE DRUG REACTIONS

There will be no reporting of adverse events/reactions during these validation studies.

12 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Reports will be submitted annually for 6 years, with final reports after completion of the sample validation after year 1 of the study period and after completion of all-case validation at the end of the study period.

In addition, dissemination activities will be undertaken including articles in scientific journals, and presentations at conferences. Publications will be developed according to UCB policies and authorship will be determined in accordance with the International Committee of Medical Journal Editors (ICMJE) guidelines.

13 REFERENCES

Aberra FN, Stettler N, Brensinger C, Lichtenstein GR, Lewis JD. Risk for active tuberculosis in inflammatory bowel disease patients. Clin Gastroenterol Hepatol. 2007;5(9):1070-5. doi:10.1016/j.cgh.2007.04.007 (published Online First: 14 Jul 2007).

Andersohn F, Schade R, Suissa S, Garbe E. Cyclooxygenase-2 selective nonsteroidal anti-inflammatory drugs and the risk of ischemic stroke: a nested case-control study. Stroke. 2006;37(7):1725-30. doi:10.1161/01.STR.0000226642.55207.94 (published Online First: 25 May 2006).

Bezin J, Girodet PO, Rambelomanana S, Touya M, Ferreira P, Gilleron V, et al. Choice of ICD-10 codes for the identification of acute coronary syndrome in the French hospitalization database. Fundam Clin Pharmacol. 2015;29(6):586-91. doi:10.1111/fcp.12143 (published Online First: 25 Aug 2015).

Cazzola M, Puxeddu E, Bettoncelli G, Novelli L, Segreti A, Cricelli C, et al. The prevalence of asthma and COPD in Italy: A practice-based study. Respir Med. 2011;105(3):386-91.

Coloma PM, Valkhoff VE, Mazzaglia G, Nielsson MS, Pedersen L, Molokhia M, et al. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. BMJ Open. 2013;3(6) doi:10.1136/bmjopen-2013-002862 (published Online First: 26 Jun 2013).

Cricelli C, Mazzaglia G, Samani F, Marchi M, Sabatini A, Nardi R, et al. Prevalence estimates for chronic diseases in Italy: Exploring the differences between self-report and primary care databases. J Public Health Med. 2003;25(3):254-7.

De Wilde S, Carey IM, Bremner SA, et al. A comparison of the recording of 30 common childhood conditions in the Doctor's Independent Network and General Practice Research Databases. Health Stat Q. 2004;2:21-30.

EUPAS7674 CA. Non-interventional Study Final Report - Multinational database cohort study to assess RMP-specified safety outcomes in association with indacaterol/glycopyrronium bromide in Europe 2019. Available from: http://www.encepp.eu/encepp/viewResource.htm;jsessionid=GfBiyQF2ir-q1UWsiKgKn_nrGAX-CUejc1rfBCoEp4d6HFYDdFJG_!846322558?id=7675. Last accessed 16 Jun 2020.

Filippi A, Vanuzzo D, Bignamini AA, Mazzaglia G, Cricelli C, Catapano AL. The database of Italian general practitioners allows a reliable determination of the prevalence of myocardial infarction. Ital Heart J. 2005;6(4):311-4.

García-Gil Mdel M, Hermosilla E, Prieto-Alhambra D, Fina F, Rosell M, Ramos R, et al. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). Inform Prim Care. 2011;19(3):135-45.

Giroud M, Hommel M, Benzenine E, Fauconnier J, Béjot Y, Quantin C, et al. Positive predictive value of French hospitalization discharge codes for stroke and transient ischemic attack. Eur Neurol. 2015;74(1-2):92-9.

Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. 2013;346:f2350. doi:10.1136/bmj.f2350 [published Online First: 23 May 2013].

Holland-Bill L, Xu H, Sørensen HT,Acquavella J, Sværke C, Gammelager H, et al. Positive predictive value of primary inpatient discharge diagnoses of infection among cancer patients in the Danish National Registry of Patients. Ann Epidemiol. 2014;24(8):593-7, 97 e1-18. doi:10.1016/j.annepidem.2014.05.011 (published Online First: 03 Aug 2014).

International Society of Pharmacoepidemiology (ISPE). Guidelines for good pharmacoepidemiology practices (GPP) (Revision 3). Jun 2015. Available from: https://www.pharmacoepi.org/resources/policies/guidelines-08027/. Accessed on 16 Jun 2020.

Jick SS, Lieberman ES, Rahman MU, Choi HK. Glucocorticoid use, other associated factors, and the risk of tuberculosis. Arthritis Rheum. 2006;55(1):19-26. doi:10.1002/art.21705 (published Online First: 08 Feb 2006).

Krarup LH, Boysen G, Janjua H, Prescott E, Truelson T. Validity of stroke diagnoses in a National Register of Patients. Neuroepidemiol. 2007;28(3):150-4. doi:10.1159/000102143 (published Online First: 05 May 2007).

Madsen M, Davidsen M, Rasmussen S, Abildstrom SZ, Osler M. The validity of the diagnosis of acute myocardial infarction in routine statistics: A comparison of mortality and hospital discharge data with the Danish MONICA registry. J Clin Epidemiol. 2003;56(2):124-30.

Mendis S, Thygesen K, Kuulasmaa K, Giampaoli S, Mähönen M, Ngu Blackett K, et al. World Health Organization definition of myocardial infarction: 2008-09 revision. Int J Epidemiol 2011;40(1):139-46. doi:10.1093/ije/dyq165 (published Online First: 12 Oct 2010).

Ohlmeier C, Langner I, Garbe E, Riedel O. Validating mortality in the German Pharmacoepidemiological Research Database (GePaRD) against a mortality registry.

Pharmacoepidemiol Drug Saf. 2016;25(7):778-84. doi:10.1002/pds.4005 (published Online First: 12 Apr 2016).

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 (published Online First: 20 Jun 2015).

Ohlmeier C, Frick J, Prütz F, Lampert T, Ziese T, Mikolajczyk R, et al. Use of routine data from statutory health insurances for federal health monitoring purposes. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2014;57(4):464-72.

Pigeot I, Ahrens W. Establishment of a pharmacoepidemiological database in Germany: Methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf. 2008;17(3):215-23.

Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, 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(7):2064-89. doi:10.1161/STR.0b013e318296aeca (published Online First: 09 May 2013).

Sahli L, Lapeyre-Mestre M, Derumeaux H, Moulis G. Positive predictive values of selected hospital discharge diagnoses to identify infections responsible for hospitalization in the French national hospital database. Pharmacoepidemiol Drug Saf. 2016;25(7):785-9. doi:10.1002/pds.4006 (published Online First: 29 Apr 2016).

Schink T, Kollhorst B, Varas Lorenzo C, Arfè A, Herings R, Lucchi S, et al. Risk of ischemic stroke and the use of individual non-steroidal anti-inflammatory drugs: A multi-country European database study within the SOS Project. PLoS One. 2018;13(9):e0203362. doi:10.1371/journal.pone.0203362 (published Online First: 20 Sep 2018).

Schink T, Holstiege J, Kowalzik F, Zepp F, Garbe E. Risk of febrile convulsions after MMRV vaccination in comparison to MMR or MMR+V vaccination. Vaccine. 2014;32(6):645-50.

Schmedt N, Kollhorst B, Enders D, Jobski K, Krappweis J, Garbe E, et al. Comparative risk of death in older adults treated with antipsychotics: A population-based cohort study. Eur Neuropsychopharmacol. 2016a;26(9):1390-400.

Schmedt N, Jobski K, Kollhorst B, Krappweis J, Rüther E, Schink T, et al. Treatment patterns and characteristics of older antipsychotic users in Germany. Int Clin Psychopharmacol. 2016b;31(3):159-69.

Sundbøll J, Adelborg K, Munch T, Frøslev T, Sørensen HT Bøtker HE, et al. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. BMJ Open. 2016;6(11):e012832. doi:10.1136/bmjopen-2016-012832 (published Online First: 20 Nov 2016).

Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow, DA, et al. Fourth universal definition of myocardial infarction. Circulation. 2018;138(20):e618-51. doi:10.1161/CIR.0000000000000617 (published Online First: 21 Dec 2018).

van Staa TP, Abenhaim L. The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycemia and other conditions. Pharmacoepidemiol Drug Saf. 1994;3:15-21.

Vlug AE, van der Lei J, Mosseveld BM, van Wijk MA, van der Linden PD, Sturkenboom MC, et al. Postmarketing surveillance based on electronic patient records: The IPCI Project. Methods Inf Med. 1999;38(4-5):339-44.

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbrouche JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9. doi: 0.1016/j.jclinepi.2007.11.008.

Voordouw AC, Sturkenboom MC, Dieleman JP, Stijnen T, Smith DJ, van der Lei J, et al. Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. JAMA. 2004;292(17):2089-95.

World Health Organization Collaborating Centre for Drug Statistics Methodology. ACT/DDD Index 2019. Available from: https://www.whocc.no/atc_ddd_index/. Accessed on 16 Jun 2020.

APPENDIX 1. LIST OF STAND-ALONE DOCUMENTS

Number	Document reference number	Date	Title
1	Appendix 1	16 Jun 2020	List of Stand-Alone Documents
2	Appendix 2	16 Jun 2020	List of Diagnosis Codes
3	Appendix 3	16 Jun 2020	Validation Algorithm: Cardiovascular Death
4	Appendix 4	16 Jun 2020	Validation Algorithm: Myocardial Infarction
5	Appendix 5	16 Jun 2020	Validation Algorithm: Stroke

APPENDIX 2. LIST OF DIAGNOSIS CODES

The following code lists provided in this appendix are preliminary and will be reviewed by all databases prior to data extraction for the CV PASS and SI PASS. As coding might change over time, relevant codes might be updated and additional codes may be added during the course of the project, if necessary.

Table A: Code list for identification of MI and stroke

Terms	ICD-10	ICD-9-CM	ICPC	ICD-10-GM
Myocardial infarction				
Cardiac infarction	I22*	--	--	I22*
Cardiac infarction	I21*	--	--	I21*
Acute myocardial infarction	I21*	410*	K75	I21*
Acute myocardial infarction, unspecified	I21.9	410.9	--	I21.9
Myocardial infarction (acute) NOS	I21.3	410	--	I21.2
Acute myocardial infarction, unspecified site, episode of care unspecified	--	410.90	--	
AMI NOS, unspecified	--	410.90	--	
Acute myocardial infarction sub endocardial infarction	--	410.7	--	
Old myocardial infarction [#]	I25.2	412		I25.2*
Healed myocardial infarction [#]	--	--	--	
Subsequent/recurrent myocardial infarction	I22	--	--	I22*
Subsequent myocardial infarction of unspecified site	I22.9	--	--	I22.9
Subsequent myocardial infarction of other sites	I22.8	--	--	I22.8
Subsequent myocardial infarction of anterior wall	I22.0	--	--	I22.0
Subsequent myocardial infarction of inferior wall	I22.1	--	--	I22.1
Subsequent acute sub endocardial myocardial infarction	I22.2	--	--	n.A.
Subsequent non transmural myocardial infarction NOS	I22.2	--	--	n.A.
Subsequent myocardial infarction (acute) NOS	I22.9	--	--	n.A.
Re-infarction of myocardium	--	--	--	
Acute sub endocardial myocardial infarction	I21.4	--	--	I21.4
Acute myocardial infarction sub endocardial infarction, episode of care unspecified	--	410.70	--	
Non transmural myocardial infarction	I21.4	--	--	n.A.
Acute myocardial infarction, of antero lateral wall	--	410.0	--	
Acute antero septal myocardial infarction	--	--	--	
Acute inferior myocardial infarction	--	410.4	--	

Table A: Code list for identification of MI and stroke

Terms	ICD-10	ICD-9-CM	ICPC	ICD-10-GM
Acute myocardial infarction, true posterior wall infarction	--	410.6	--	
True posterior myocardial infarction	--	--	--	
Acute myocardial infarction, of inferoposterior wall	--	410.3	--	
Other specified anterior myocardial infarction	--	--	--	
Acute transmural myocardial infarction of unspecified site	I21.3	--	--	I21.3
Acute transmural myocardial infarction of anterior wall	I21.0, 122.0	--	--	I21.0
Acute transmural myocardial infarction of inferior wall	I21.1, I21.19, 122.1	--	--	I21.1
Acute transmural myocardial infarction of other sites	I21.2, I21.29, 122.8	--	--	I21.2, I21.3,
ECG: old myocardial infarction #	--	--	--	
Anterior myocard. infarct NOS	--	410.8	--	
Other acute myocardial infarct	--	--	--	
Other acute myocardial inf. NOS	--	--	--	
Inferior myocard. infarct NOS	--	--	--	
Acute myocardial infarction, of infero lateral wall	--	410.2	--	
Acute lateral myocardial infarction	--	410.5	--	
Lateral myocardial infarct NOS	--	--	--	
Acute widespread myocardial infarction	--	--	--	
Acute posterior myocardial infarction	--	410.60, 410.61, 401.62	--	
Posterior myocard. infarct NOS	--	--	--	
Silent myocardial infarct #	--	--	--	
ECG: myocardial infarction	--	--	--	
ECG: myocardial infarct NOS	--	--	--	
Postoperative sub endocardial myocardial infarction	--	--	--	
Postoperative myocardial infarction	--	--	--	
Acute anterior myocardial infarction	--	410.1	--	
Acute Q wave myocard infarct	--	--	--	
Acute myocardial infarction sub endocardial infarction	--	410.71, 410.72	--	
Non-Q wave myocardial infarction NOS	I21.4, I22.2	--	--	I21.4
Non-ST elevation (NSTEMI) myocardial infarction	I21.4, I22.2	--	--	I21.4

Table A: Code list for identification of MI and stroke

Terms	ICD-10	ICD-9-CM	ICPC	ICD-10-GM
History of MI [#]	--	--	K76.02	I25.2*
Diabetes mellitus insulin-glucose infuse acute myocardial infarct	--	--	--	
Stroke				
Terms	**ICD10**	**ICD-9-CM**	**ICPC**	**ICD-10-GM**
Stroke, not specified as hemorrhage or	I64	--	--	I64
Stroke NOS	I63.9	--	K90	I63.9
Non-traumatic subarachnoidal bleeding	I60	430	--	I60*
Intracerebral haemorrhage	I61	431	--	I61*
Cerebrovascular accident (CVA)	--	--	--	
Stroke and cerebrovascular accident unspecified	--	--	--	
Stroke NOS	--	--	--	
Sequelae of stroke, not specified as hemorrhage or infarction [b]	I69	342	--	I69*
Brain stem stroke syndrome	G46.3	--	--	G46.3
Cerebellar stroke syndrome	G46.4	--	--	G46.4
Other and unspecified intracranial haemorrhage	I62	432.*	--	I62*
Cerebral infarction	I63	--	--	I63*
Personal history of stroke [#]	--	--	--	
Sequelae of stroke NOS [#]	I69.3	--	--	I69.3
H/O: Stroke [§]	--	--	--	
Cerebral infarct due to thrombosis of precerebral arteries	--	433*	--	
Personal history of transient ischemic attack (TIA), and cerebral infarction without residual deficits [#]	Z86.73	V12.54	--	Z86.7
Management/monitoring of stroke	--	--	--	
Delivery of rehabilitation for stroke	--	--	--	
Stroke referral, Seen in stroke clinic	--	--	--	
Quality indicators stroke	--	--	--	
Sequelae of cerebral infarction	--	--	--	

Table A: Code list for identification of MI and stroke

Terms	ICD-10	ICD-9-CM	ICPC	ICD-10-GM
Sequelae of stroke, not specified as haemorrhage or infarction [#]	--	438.*	--	
Cerebral infarction due unspecified occlusion/stenosis of pre-cerebral arteries	--	433.*	--	
Cerebral infarction due to unspecified occlusion/stenosis of cerebral arteries		434.*		
[X] Other cerebral infarction	--	--	--	
[X] Occlusion and stenosis of other cerebral arteries	--	--	--	
Discharge from stroke service	--	--	--	
Hemiplegia and hemiparesis	--	342.*	--	
Acute, but ill-defined, cerebrovascular disease	--	436.*	--	

AMI=acute myocardial infarction; ECG=electrocardiogram; H/O=history of; ICD-9-CM=International Classification of Diseases, revision 9, clinical modification; ICD-10=International Classification of Diseases, revision 10; ICD-10-GM=International Classification of Diseases, revision 10, German modification; ICPC=International Classification of Primary Care; MI=myocardial infarction; n.A.=not applicable; NOS=not otherwise specified

* Includes subcodes.

[#] Not for an acute event; will only be considered for stroke as underlying comorbidity.

[§] History of. Not to be used for the identification of incidence events.

The following code list in Table B should be considered preliminary for this study and will be reviewed and subsequently enlarged by all databases prior to data extraction for the SI PASS.

Table B: Diagnosis codes used in case-ascertainment algorithms for infections (ICD-10 codes by Holland-Bill et al, 2014)

Infection					
Diagnosis	**ICD-10 Denmark**	**ICD-10**	**ICD-9-CM**	**ICPC**	**ICD-10-GM**
Pneumonia	J12xx - J18xx	J12xx - J18xx	480* - 487*	R81	J12xx - J18xx
Sepsis	A40xx - A41xx, B377, A327, A548G, A021, A227, A267, A427, A282B	A40xx - A41xx, B37.7, A32.7, A54.8, A02.1, A22.7, A26.7, A42.7, A28.2	003.1, 020.2, 022.3, 027.0, 027.1, 036.2, 038*, 054.5, 519.0, 599.0, 639.0, 659.3, 771.0*, 998.59, 999.3	Free text – no disease code IPCE	A40xx - A41xx, B377, A327, A548, A021, A227, A267, A427, A282
Skin infection	H010x, H03xx, H600x, H601x, H602x, H603x, H62xx, K122x, K130x, K61xx, L01xx, L08xx, M726x, A46xx	H01.0x, (H00.03) H60.0x, H60.1x, H60.2x, H60.3x, H62xx, K12.2x, K13.0x, K61xx, L01xx, L08xx, M72.6x, A46xx	372.2*, 373.0*, 373.6*, 380.0*, 380.1*, 380.2*, 112.82, 054.73, 054.71, 528.5*, 566, 680*, 684*, 686*, 039.0, 728.86, 035	S76 S10.01 S10.02	H010x, H03xx, H600x, H601x, H602x, H603x, H62xx, K122x, K130x, K61xx, L01xx, L08xx, M726x, A46xx
Cellulitis	L03xx	L03xx	670*, 614.4, 478.5 528.3, 528.5, 597.0 376.01 475	S10.03	L03xx

Table B: Diagnosis codes used in case-ascertainment algorithms for infections (ICD-10 codes by Holland-Bill et al, 2014)

Infection					
Diagnosis	**ICD-10 Denmark**	**ICD-10**	**ICD-9-CM**	**ICPC**	**ICD-10-GM**
			607.2, 566, 608.4, 478.21 478.71, 614.4, 614.3		
Endocarditis	I33xx, I398, B376	I33xx, I39, B37.6	093.2*, 098.84 112.81, 424.9* 036.42, 115.04, 115.14, 115.94, 074.22, 421.0, 421.1	K71	I33xx, I398, B376
Bacteremia	A499	A49.9		Free text – no disease code IPCI	A499
Meningitis	G00xx, G01xx, G02xx, G03xx, A321, A390, A170x, A203, A87x, A548D, A022C, B375, B003, B010, B021, B051, B261, B384	G00xx, G01xx, G02xx, G03xx, A32.1, A39.0, A17.0x, A20.3, A87x, A54.8D, A02.2, B37.5, B00.3, B01.0, B02.1, B05.1, B26.1, B38.4	049.1, 098.82, 047*, 090.42, 112.83, 320*, 321*, 322*, 036.0, 094.2, 114.2, 115.01, 036.0, 115.11, 115.91, 054.72, 100.81, 003.21, 091.81, 013.0*	N71.01 N71.02	G00xx, G01xx, G02xx, G03xx, A321, A390, A170x, A203, A87x, A548, A022, B375, B003, B010, B021, B051, B261, B384
Encephalitis, myelitis,	G04xx, G05xx	G04xx, G05xx	348.3, 323.0,	N71.03 N71.04	G04xx, G05xx

Table B: Diagnosis codes used in case-ascertainment algorithms for infections (ICD-10 codes by Holland-Bill et al, 2014)

Infection					
Diagnosis	**ICD-10 Denmark**	**ICD-10**	**ICD-9-CM**	**ICPC**	**ICD-10-GM**
encephalomyelitis			323.1, 323.2, 323.5, 323.6, 323.7, 049.8, 049.9, 052.0, 055.0, 090.41, 130.0, 056.01, 094.81, V73.5, 036.1, 054.3, 064*, 063.2, 066.2, 323.2, 323.1, 323.0, 063* 013.6*, 072.2, 341.1, 323.5, 063.8, 062*, 487.8, 330.8, 046.1		
Urinary tract infection	N10xx, N11xx, N12xx, N151x, N159x, N30xx, N34xx, N390x	N10xx, N11xx, N12xx, N15.1x, N15.9x, N30xx, N34xx, N39.0x	583*, 590*, 595*, 599.0, 580.9*, 597*	U71	N10xx, N11xx, N12xx, N151x, N159x, N30xx, N34xx, N390x
Kidney infection (pyelonephritis)	N10xx, N11xx, N12xx, N151x, N159x	N10xx, N11xx, N12xx, N15.1x, N15.9x	583*, 590*, 580.9*	U70	N10xx, N11xx, N12xx, N151x, N159x

Table B: Diagnosis codes used in case-ascertainment algorithms for infections (ICD-10 codes by Holland-Bill et al, 2014)

Infection					
Diagnosis	**ICD-10 Denmark**	**ICD-10**	**ICD-9-CM**	**ICPC**	**ICD-10-GM**
Septic arthritis	M00xx	M00xx	711.0*	L70.02	M00xx
Infective arthritis	M011x, M013x	M01.X1, M01.X3	711.1* - 19*	L70.02	M011x, M013x
Osteomyelitis	M86xx	M86xx	730.0*, 730.1*, 730.2*	L70.01	M86xx
Tuberculosis	A15xx - A19xx	A15xx - A19xx	010* - 018*, V12.01	R70	A15xx - A19xx
Atypical mycobacteria	A31xx	A31xx	031*	No code in IPCI	A31xx
Mycoses	B35xx - B49xx	B35xx - B49xx	110*-118*, 711.6*	S74	B35xx - B49xx
Systemic candidiasis	B37xx	B37xx	112*	S75	B37xx
Cryptococcosis	B45xx	B45xx	117.5*	No code in IPCI	B45xx
Aspergillosis	B44xx	B44xx	117.3, 484.6	No code in IPCI	B44xx
Histoplasmosis	B39xx	B39xx	115*	No code in IPCI	B39xx

ICD-9-CM=International Classification of Diseases, revision 9, clinical modification; ICD-10=International Classification of Diseases, revision 10; ICD-10-GM=International Classification of Diseases, revision 10, German modification; ICPC=International Classification of Primary Care; IPCI=Integrated Primary Care Information Project

Table C: Code list for the identification of CV events for validation of CV death

CV events		
Diagnosis	**ICD-10**	**ICD-10-GM**
Heart failure	I50, I50.0, I50.1, I50.9, I11.0, I13.0, I13.2	I500x, I501x, I509, I110, I1100, I1101, I130, I1300, I1301, I132, I1320, I1321
Sudden cardiac death	I46, I46.1	I461
Cardiac arrhythmia	I44, I44.0, I44.1, I44.2, I44.3, I45.6, I45.81, I47, I47.0, I47.1, I47.2, I47.9, I48, I48.1, I48.2, I48.3, I48.4, I48.9, I48.92, I49.0, I49.1, I49.2, I49.3, I49.4, I49.49, I49.5,	I440, I441, I442, I443, I444, I445, I446, I447, I450, I451, I452, I453, I454, I455, I456, I458, I459, I480x, I481x, I482, I483, I484, I489, I490, I491, I492, I493, I494, I495, I498, I499

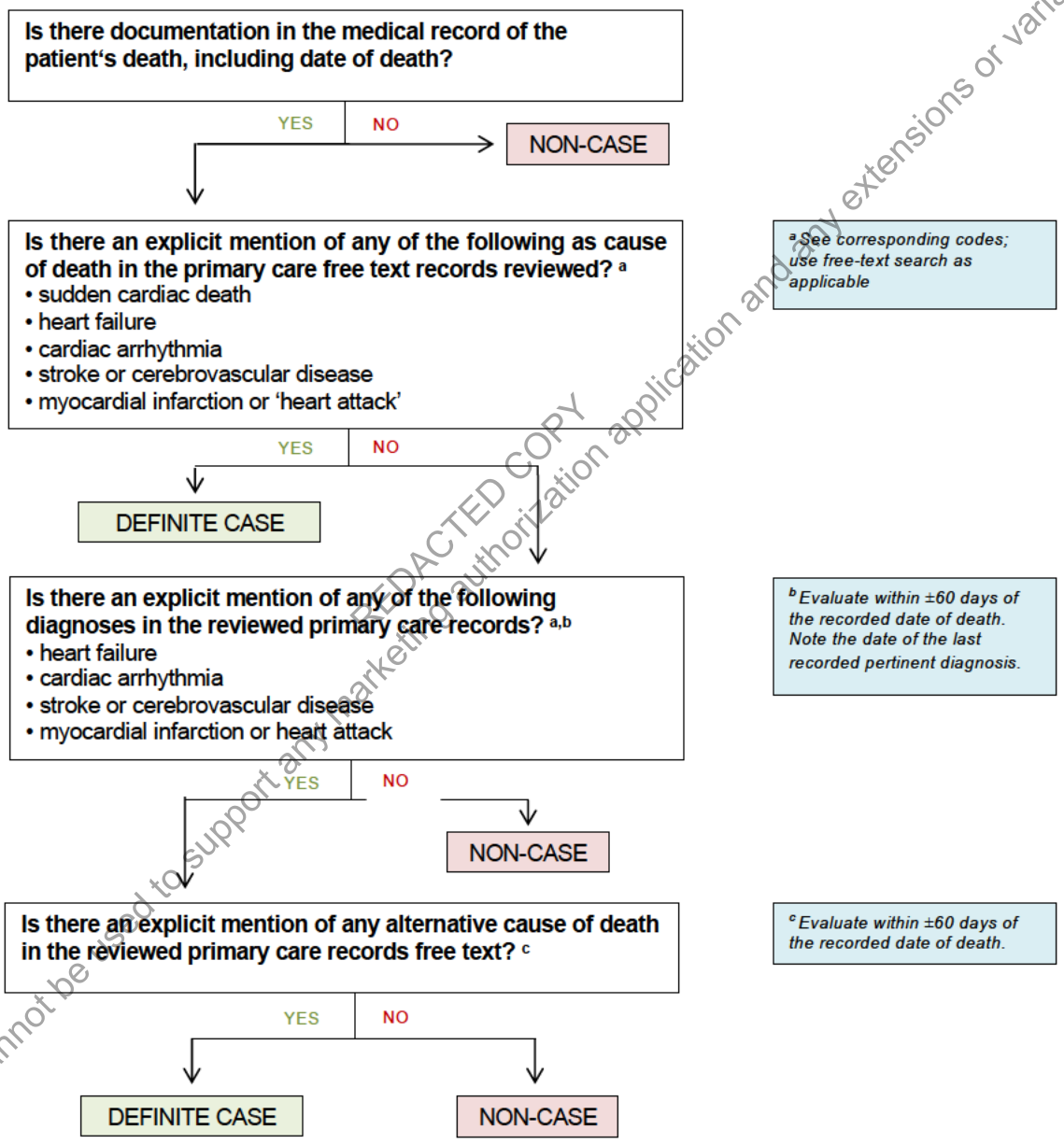
CV=cardiovascular; HSD=Health Search Database; ICD-10=International Classification of Diseases, revision 10; ICD-10-GM=International Classification of Diseases, revision 10, German modification; IPCI=Integrated Primary Care Information Project

Note: For IPCI and HSD, CV events around the time of death will be identified using free text review.

APPENDIX 3. VALIDATION ALGORITHM: CARDIOVASCULAR DEATH

The following algorithm has been used in a previous EU-ADR validation study.

Note: If insufficient data to assess, label case as "**Not Assessable**".



Questionnaire for assessors

Based on the information reported in the clinical definition, assessors may use the following questionnaire on a voluntary basis to be able to apply subsequently a validation algorithm:

Database: _____
Patient ID: _____
Gender: M/F
Birthdate: (dd/mm/yyyy) _____
Date of event as reported in Jerboa output: (dd/mm/yyyy) _____

Endpoint name CV death "DEFINITE CASE", "NON-CASE", "NON-ASSESSABLE"
Date of death upon validation (dd/mm/yyyy): _____

1.) Is there documentation in the medical record of the patient's death, including date of death?
If YES, then NON-CASE,
If NO, proceed to next question below:

2.) Is there an explicit mention of any of the following as cause of death in the primary care free text records reviewed?
- sudden cardiac death
- heart failure
- cardiac arrhythmia
- myocardial infarction (or 'heart attack')
- stroke or cerebrovascular disease

If YES, then DEFINITE CASE (see algorithm).
If NO, proceed to next question below:

3.) Is there an explicit mention of any of the following diagnoses in the reviewed primary care records?
- heart failure
- cardiac arrhythmia
- stroke or cerebrovascular disease
- myocardial infarction (or 'heart attack')

Evaluate within ±60 days of the recorded date of death.

Note the date of the last recorded pertinent diagnosis.

If NO, then NON-CASE.
If YES, then proceed to next question below:

4.) Is there an explicit mention of any alternative cause of death in the reviewed primary care records free text?

If YES, then NON-CASE.

If NO, then DEFINITE CASE.

Evaluate within ±60 days of the recorded date of death.

Comments:

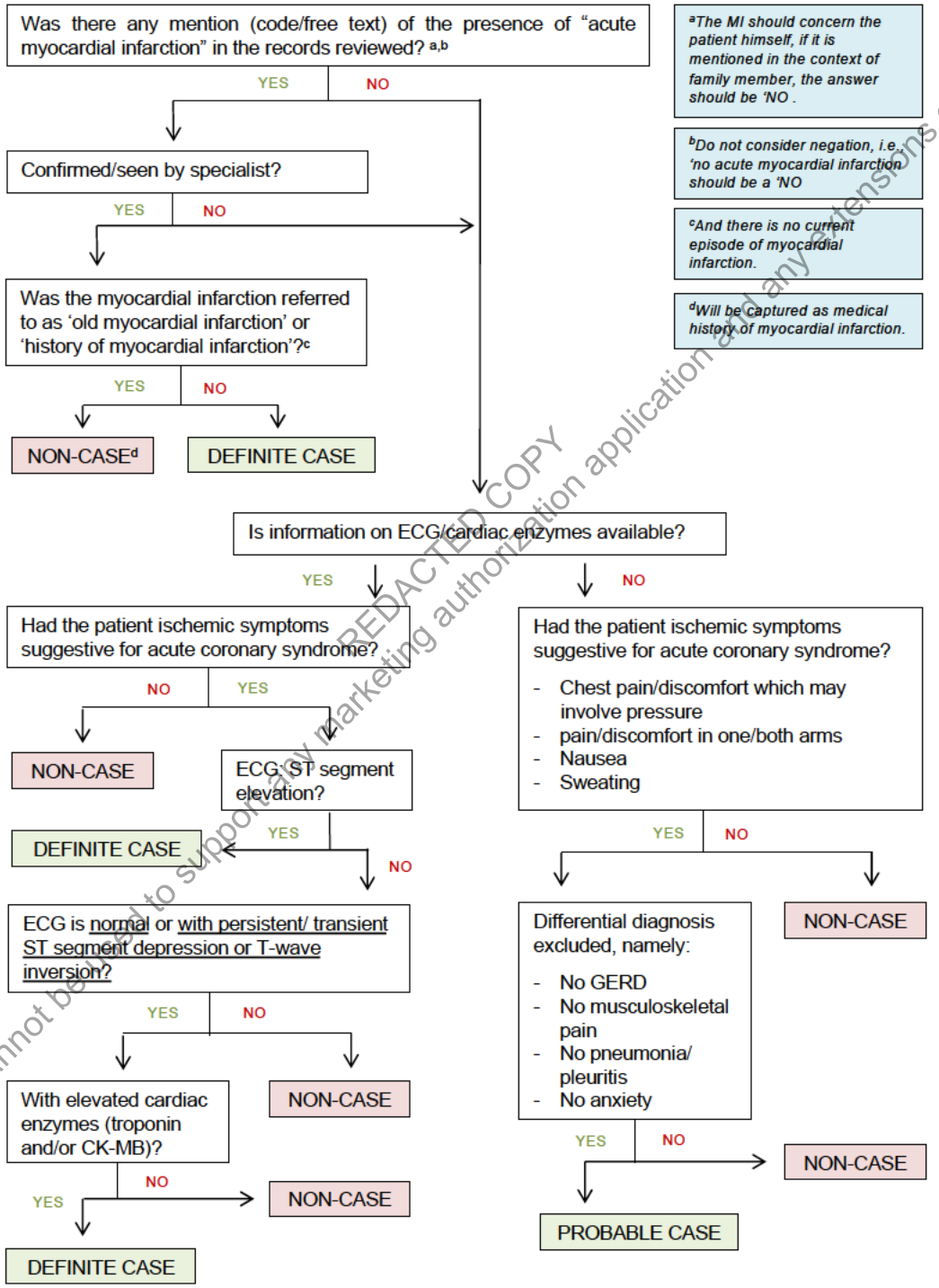
_____

_____

APPENDIX 4. VALIDATION ALGORITHM: MYOCARDIAL INFARCTION

The algorithm was used for case validation in a previous EU-ADR validation study for MI.

Note: If insufficient data to assess, label case as "**Not Assessable**".



Questionnaire for assessors

Based on the information reported in the clinical definition, assessors may use the following questionnaire on a voluntary basis to be able to apply subsequently a validation algorithm:

Database: _____
Patient ID: _____
Gender: M/F
Birthdate: (dd/mm/yyyy) _____
Date of event as reported in Jerboa output: (dd/mm/yyyy) _____

Endpoint name MI "DEFINITE CASE", "PROBABLE CASE", "NON-CASE", "NON-ASSESSABLE"
Date of MI upon validation (dd/mm/yyyy): _____

1.) Was there any mention (code/free text) of the presence of "acute myocardial infarction" in the records reviewed?[a,b]
    If YES, then proceed to question (2) below:
    If NO, then proceed to question (4) below:

[a]*The AMI should concern the patient himself, if it is mentioned in the context of family member, the answer should be NO.*
[b]*Do not consider negation: "No acute myocardial infarction should be NO.*

2.) Confirmed/seen by a specialist?
    If YES, then proceed to question (3) below:
    If NO, then proceed to question (4) below:

3.) Was the myocardial infarction referred to as 'old myocardial infarction' or 'history of myocardial infarction'?[c]
    If YES, then NON-CASE
    If NO, then DEFINITE CASE

[c]*And there is no current episode of myocardial infarction?*

4.) Is information on ECG/cardiac enzymes available?
    If YES, then proceed to question (5) below
    If NO, then proceed to question (9) below:

5.) Had the patient ischemic symptoms suggestive for acute coronary syndrome?
    If YES, then proceed to question (6) below:
    IF NO, then NON-CASE

6.) Was the ST segment elevated in the ECG?
    If YES, DEFINITE CASE
    If NO, then proceed to question (7) below:

7.) Was the ECG normal or with persistent/transient ST segment depression or T-wave inversion?
If YES, then proceed to question (8) below:
If No, NON-CASE

8.) Were elevated cardiac enzymes recorded (troponin and/or CK-MB)?
If YES, DEFINITE CASE
If NO, NON-CASE

9.) Had the patient ischemic symptoms suggestive for acute coronary syndrome (chest pain/discomfort which may involve pressure; pain/discomfort in one/both arms; nausea; sweating) and no ECG/cardiac enzyme test results are available?
If YES, proceed to question (10) below:
If NO, NON-CASE

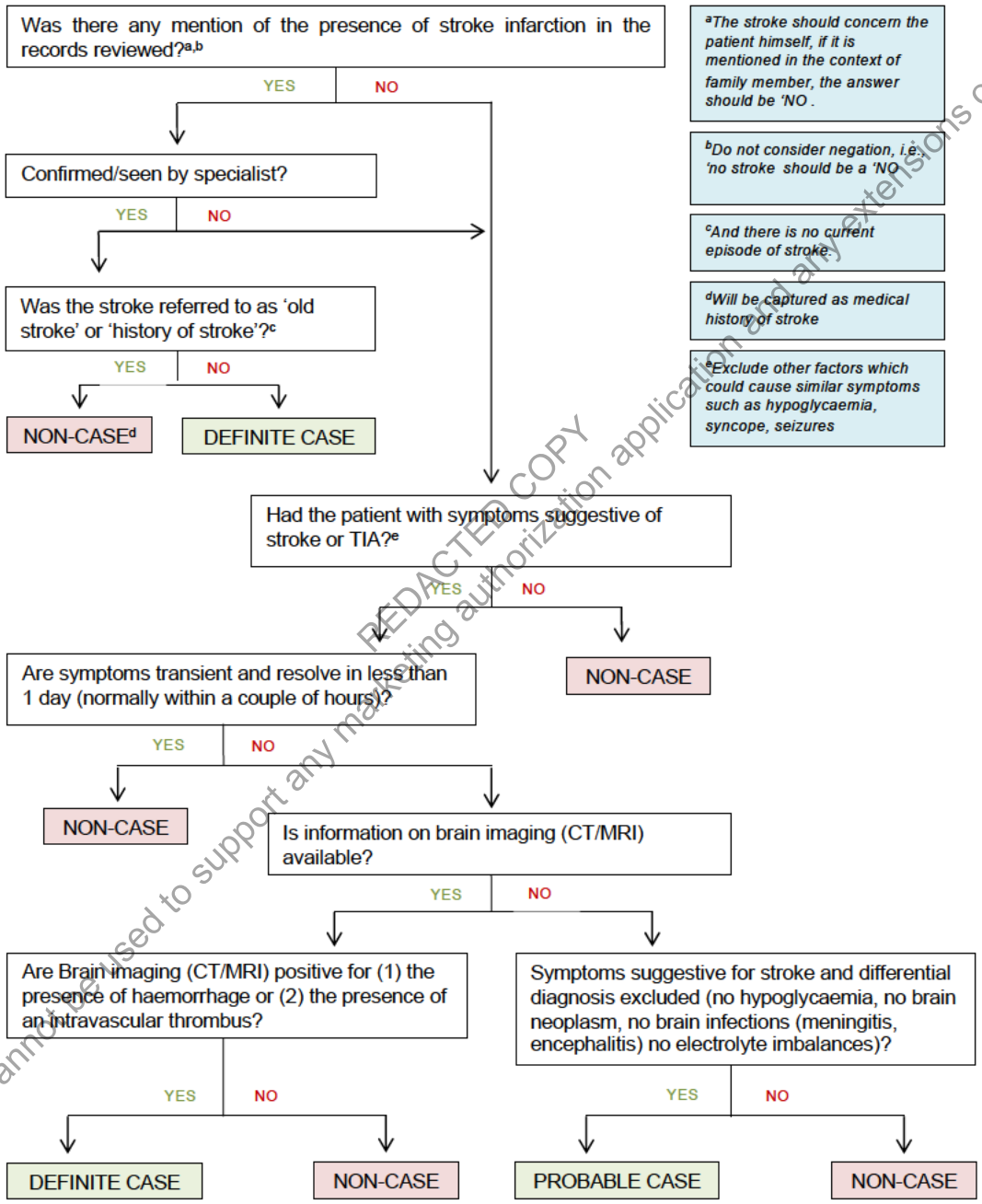
10.) Were differential diagnoses, namely GERD, musculoskeletal pain, pneumonia/pleuritis and anxiety excluded?
If YES, PROBABLE CASE
If NO, NON-CASE

Comments:
_____
_____

APPENDIX 5. VALIDATION ALGORITHM: STROKE

The following algorithm has been used in a previous EU-ADR validation study.

Note: If insufficient data to assess, label case as **"Not Assessable".**



Questionnaire for assessors

Based on the information reported in the clinical definition, assessors may use the following questionnaire on a voluntary basis to be able to apply subsequently a validation algorithm.

Database: _____
Patient ID: _____
Gender: M/F
Birthdate: (dd/mm/yyyy)_____

Date of event as reported in Jerboa output: (dd/mm/yyyy) Validation of stroke:

Endpoint name stroke "DEFINITE CASE", "PROBABLE CASE", "NON-CASE", "NON-ASSESSABLE"
Date of stroke upon validation (dd/mm/yyyy): _____

1.) Was there any mention of the presence of stroke infarction in the reviewed records?
    If YES, proceed to question 2
    If NO, proceed to question 3.

2.) Confirmed/seen by a specialist?
    If YES, proceed to question 3
    If NO, proceed to question 4

3.) Was the stroke referred to as 'old stroke' or 'history of stroke'?
    If YES, then NON-CASE
    If NO, DEFINITE CASE

4.) Had the patient with symptoms suggestive of stroke or TIA?
    If YES, proceed to question 5
    If NO, NON-CASE

5.) Are symptoms transient and resolve in less than 1 day (normally within a couple of hours)?'
    If YES, then NON-CASE.
    If NO, then proceed to question 6

6.) Is information on brain imaging available?
    If YES, proceed to question 7
    If NO, proceed to question 8

7.) Are Brain imaging (CT/MRI) positive for (1) the presence of haemorrhage or (2) the presence of an intravascular thrombus?

If YES, DEFINITE CASE

If NO, NON-CASE

8.) Were symptoms suggestive for stroke and differential diagnoses excluded (e.g. no hypoglycaemia, no brain neoplasm, no brain infections (meningitis, encephalitis), no electrolyte imbalance)?

If YES, PROBABLE CASE

If NO, NON-CASE

Comments:

Approval Signatures

Name: op0007-protocol

Version: 1.0

Document Number: CLIN-000153040

Title: OP0007 Protocol - Noninterventional Validation Study of Outcomes of Interest for OP0004 and OP0006

Approved Date: 19 Jun 2020

Document Approvals	
Approval Verdict: Approved	Name: [redacted] Capacity: Subject Matter Expert Date of Signature: 18-Jun-2020 23:50:23 GMT+0000
Approval Verdict: Approved	Name: [redacted] Capacity: Medical Date of Signature: 19-Jun-2020 04:42:14 GMT+0000