BMJ Open

Characterisation of a clinical trial–like population of high cardiovascular risk patients prior to myocardial infarction or stroke in the real world: design and protocol for a multidatabase retrospective cohort study

Andreas Ochs [1], Queenie Chan [1], Nafeesa N Dhalwani,[2] Michael Duxbury [1], Erin Shannon [2], James O'Kelly [1], Gabriel Paiva da Silva Lima [3], Suna Avcil,[3] Adrienne YL Chan,[4] Celine SL Chui [5,6,7], Edward Chia-Cheng Lai [8,9], Thomas Cars [10], Ju-Young (Judy) Shin [11], Edith M Heintjes [12], Ian CK Wong[4,7,13]

To cite: Ochs A, Chan Q, Dhalwani NN, *et al*. Characterisation of a clinical trial–like population of high cardiovascular risk patients prior to myocardial infarction or stroke in the real world: design and protocol for a multidatabase retrospective cohort study. *BMJ Open* 2025;**15**:e092702. doi:10.1136/bmjopen-2024-092702

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-092702).

Received 20 August 2024
Accepted 23 June 2025

For numbered affiliations see end of article.

Correspondence to
Dr Andreas Ochs;
aochs@amgen.com

ABSTRACT

Introduction Cardiovascular (CV) disease is the leading cause of morbidity and mortality globally. Low-density lipoprotein cholesterol (LDL-C) is an important modifiable risk factor of major adverse cardiovascular events. Patients without prior myocardial infarction (MI) or stroke but with established risk factors and elevated LDL-C may benefit from intensive lipid-lowering therapy (LLT); however, the size and potential healthcare burden of this population globally are not known. The benefits of evolocumab, a proprotein convertase subtilisin/kexin type 9 inhibitor, in these patients, are currently being studied in the phase 3 Effect of Evolocumab in Patients at High Cardiovascular Risk Without Prior Myocardial Infarction or Stroke (VESALIUS-CV) trial. To characterise the high-risk pre–CV-event (VESALIUS-CV–like) individuals in the real world, an observational study is being conducted across multiple countries.

Methods and analysis This retrospective cohort study will use a common protocol and an analytical common data model approach to characterise VESALIUS-CV–like individuals in the real world across different geographical regions and healthcare settings. The study period will be from 2010 to 2022, subject to data availability in study sites. Patients aged 50 years and older at high risk of CV disease but without prior MI or stroke will be included in this study. VESALIUS-CV–like individuals are defined through a combination of the following: (1) one diagnosis of coronary artery disease, cerebrovascular disease, peripheral artery disease or diabetes with microvascular complications or chronic insulin use; (2) an elevated LDL-C measurement and (3) other high-risk factors. The objectives of this study are to estimate the prevalence of VESALIUS-CV–like individuals, describe their characteristics and care pathways and estimate their incidence rates of CV events and healthcare costs. The prevalence of VESALIUS-CV–like individuals will be expressed as annual prevalence; patient characteristics at index date will be presented using summary statistics; care pathways will be summarised as LLT prescription across time; and the incidence of defined CV events will be expressed as events per person-years as well as at certain time periods. Healthcare costs will be presented as CV-related costs in different time periods.

Ethics and dissemination Approvals of the study protocol were obtained from relevant local ethics and regulatory frameworks for each participating database. The results of the study will be submitted to peer-reviewed scientific publications and presented at scientific conferences.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This study will involve countries/regions across Asia, North America and Europe.
- ⇒ A common protocol and analytical common data model are used to ensure consistency in data analysis and validity in cross-geographical comparisons.
- ⇒ Clinical trial criteria will be emulated but not exactly matched, due to differences in the real-world capture of these criteria.
- ⇒ Most databases in this study include data on care provided in public and private healthcare settings, with only a few capturing just the predominant setting, resulting in only small amounts of missed care records.

INTRODUCTION

Cardiovascular (CV) disease is the leading cause of morbidity and mortality around the world.[1] Low-density lipoprotein cholesterol (LDL-C) is an important modifiable risk factor of CV disease.[2–4] Numerous studies have consistently shown that lowering LDL-C reduces CV risk.[5] Lipid-lowering therapy (LLT) with statins has been shown to reduce LDL-C and major adverse cardiovascular events (MACE) in both primary and secondary prevention populations.[6] Non-statin LLT with evolocumab, a proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitor, has been shown to reduce the risk of CV events in a secondary prevention population with established atherosclerotic CV disease and prior myocardial infarction (MI) or stroke.[7] Currently, a global phase 3 trial (effect of evolocumab in patients at high CV risk without prior MI or stroke [VESALIUS-CV]) is underway to evaluate the effect of evolocumab on the risk of major CV events in patients at high CV risk but without a prior major ischaemic event.[8]

Here, we describe the design, including challenges in emulating clinical trial criteria in a multicountry real-world study and the use of a unified protocol and analytical common data model (ACDM) to generate comparable estimates, of this multicountry, multidatabase observational study (Global Burden Study of High-Risk [VESALIUS-CV–like] Population in the REAL-World [VESALIUS-REAL]).

Hypothesis and objectives

VESALIUS-REAL is an estimation study with no formal testing of hypothesis. The aim of this study is to characterise real-world, VESALIUS-CV trial–like, high-risk pre–CV-event individuals (VESALIUS-CV–like individuals) and examine their care pathways, healthcare burden and incidence of MACE across different countries and regions.

Primary objectives are as follows:

- To estimate the prevalence of VESALIUS-CV–like population in the real world (objective 1)
- To describe the characteristics of VESALIUS-CV–like individuals (objective 2)
- To describe the care pathways of these VESALIUS-CV–like individuals (objective 3)
- To estimate the incidence rates of CV events among VESALIUS-CV–like individuals (objective 4)

The exploratory objective is as follows:

- Estimate the healthcare costs (in US data only) of this VESALIUS-CV–like population in the real world (objective 5)

METHODS AND ANALYSIS

This is a descriptive statistical study for which data analysis began in the third quarter of 2024. The study will end in the second quarter of 2025.

Study design

This is a retrospective cohort study using healthcare databases from nine different countries and territories. The study schema is shown in figure 1. Different cohorts will be created to address the study objectives. For objective 1, annual cohorts of patients at risk of being VESALIUS-CV–like individuals will be created. Objectives 2–5 will use a cohort of VESALIUS-CV–like individuals who meet the eligibility criteria during the study period in each database.

To enable consistent analysis and reporting across the different databases in the federated data network, a common protocol, global and database-specific statistical analysis plans and an ACDM will be created to execute the study. The concept of the ACDM is to standardise a limited set of extracted variables into a common data structure, allowing the use of common analytics and methods across multiple datasets.[9] This allows for the use of unified methodology, definitions, calculations and standardisation, which facilitates cross-geographical comparisons. Each database owner will be asked to convert their raw data into the ACDM format, apply the common statistical programming code to perform the analysis locally and only share aggregated results for collation of the overall study results. No individual patient-level data will be shared.

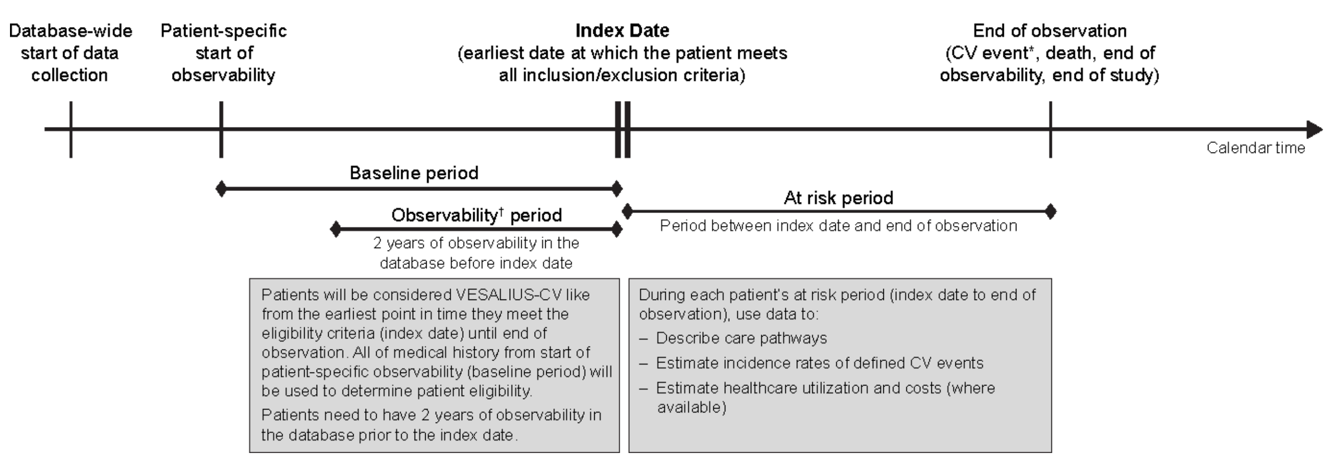

Figure 1 Study schema. Study-wide data availability: 1 January 2008 (or earliest data available) to 31 December 2022 (or at least data available). Study period: 1 January 2010 (earliest possible index date) to 31 December 2022 (or latest data available). *CV events are not considered end of observation events when survival endpoints are assessed. †Particulars of observability to be defined for each data source. CV, cardiovascular; VESALIUS-CV, Effect of Evolocumab in Patients at High Cardiovascular Risk Without Prior Myocardial Infarction or Stroke.

Data source

All included databases will use patient-level healthcare data derived from electronic medical records or claims data from nine countries and territories across three continents. The study period will be from 1 January 2010 to 31 December 2022, subject to data availability in each database. Database details are provided in tables 1 and 2 and online supplemental table 1.

A feasibility questionnaire was used to identify datapoints available in the data sources (online supplemental table 2). All databases have been assessed to be capable of contributing data for objectives 1–4. Each database will contribute data on diagnosis, procedures, laboratory values, prescriptions and death. The database from the USA will additionally contribute data for exploratory objective 5 as it is the largest database in the study. For databases where the denominator cannot be derived from the data, published national data will be used to compute prevalence estimates of objective 1.

Study population, eligibility criteria and index date

Emulation of clinical trial eligibility criteria in real-world studies can pose challenges due to clinical trial criteria being more specific compared with data captured in routine clinical practice and real-world datasets.[10] Each VESALIUS-CV eligibility criterion was translated into a set of criteria that are available in all study databases (table 3, online supplemental table 3). Translations fell into one of three categories. First, a number of criteria could be translated directly from clinical trial to real-world databases without changes. These include diagnoses for endpoints such as MI or stroke or high-risk patient identification through age variable (>65 years of age). Second, a number of criteria required generalisation of clinical trial criteria into criteria that are captured across all databases in the study. These include, for example, the generalisation of peripheral artery disease (PAD) defined in VESALIUS-CV by percentage of limb artery stenosis, abdominal aorta treatment or ankle-brachial index (ABI) by translation into PAD diagnosis codes. Cerebrovascular disease (CeVD), defined in VESALIUS-CV as prior transient ischaemic attack (TIA) with ≥50% carotid stenosis or internal or external carotid artery stenosis of ≥70% or ≥2 of ≥50% stenoses or prior internal or external carotid artery revascularisation, was generalised by translation into diagnosis and procedure codes for TIA, stenosis of carotid artery or cerebral artery syndrome. Diagnoses of familial hypercholesterolaemia (FH) are often not recorded directly in real-world databases, to be able to capture patients with FH in this study, patients will be included when they are considered to be definitive familial hypercholesterolaemia patients based on medical codes or probably FH patients based on their score in the modified Dutch Lipid Clinic Network (DCLN)[11] excluding DNA analysis or family history from the scoring system. Not all laboratory measurements used to identify patients in VESALIUS-CV are routinely available from all databases included in this study, therefore proxy diagnoses were used. For example, estimated glomerular filtration rate (eGFR) of 15 to <45 mL/min/1.73 m^2 was used in

Table 1 Databases in this study

Country/region	Database	Data nature and healthcare setting	Study period
North America			
USA	HealthVerity closed and open claims data from commercial, Medicare and Medicaid FFS claims linked to laboratory and mortality data	Claims data linked to laboratory and mortality data	2016–2023
Europe			
Germany	GDA	Electronic medical records	1992–2022
Sweden	National registries and regional data registries	Electronic medical records and healthcare registries	2010–2024
The Netherlands	Pharmo Data Network	Electronic medical records	2014–2022
United Kingdom	CPRD Aurum linked to HES and ONS	Electronic medical records	1990–2022
Asia-Pacific			
Japan	DeSC database	Claims data	2014–2023
Hong Kong	Hospital Authority	Electronic medical records	2004–2020
South Korea	NHIS-HEALS	National claims	2002–2019
Taiwan	CGRD	National claims	2010–2022

Online supplemental table 1 shows additional data points available from each data source.
CGRD, Chang Gung Research Database; CPRD, Clinical Practice Research Datalink; FFS, fee-for-service; GDA, German Disease Analyzer; HES, Hospital Episode Statistics; NHIS-HEALS, National Health Insurance Service National Health Screening; ONS, Office of National Statistics.

Table 2 Sample size estimations

Country/region	Database	Number of patients in database	Percentage of population covered	Denominator	Number of patients≥50 years of age	Percentage lipid measurements up to 2016
North America						
USA	HealthVerity closed and open claims data from commercial, Medicare and Medicaid FFS claims linked to laboratory and mortality data	156 600 000	47.2%	Number of individuals in the database	69 300 000 (age 45+ years)	26 100 000 (38%)
Europe						
Germany	GDA	4 400 000	5.3%	Number of individuals in the database	1 306 656	214 775 (16%)
Sweden	National registries and regional data registries	6 300 000	60.0%	Published population data	2 300 000	1 610 000 (70%)
The Netherlands	Pharmo Data Network	1 700 000	10%	Number of individuals in cross-section of GP, out-patient pharmacy and hospital databases	730 000	291 000 (40%)
United Kingdom	CPRD Aurum linked to HES*	41 200 000	19.8%	Number of individuals in the database	1 408 838	467 714 (33%)
Asia-Pacific						
Japan	DeSC database	10 473 339	8.4%	Number of individuals in the database	5 865 478	1 438 517 (14%)
Hong Kong	Hospital Authority	11 000 000	90.0%	Number of individuals in the database	102 405 for one district (HK West)	102 405 (100%)
South Korea	NHIS-HEALS	514 000	10.0%	Number of individuals in the database	277 885 (aged 50–79 at baseline)	277 885 (100%)
Taiwan	CGRD	1 540 000	10.0%	Number of individuals in the database	373 493 (2015, aged 65+)	373 493 (100%)

*The feasibility analysis was only conducted in CPRD Gold due to limited data access to CPRD Aurum and HES data at the time of feasibility assessment, numbers are expected to be comparable in CPRD Aurum.
CGRD, Chang Gung Research Database; CPRD, Clinical Practice Research Datalink; FFS, fee-for-service; GDA, German Disease Analyzer; GP, General Practicioner; HES, Hospital Episode Statistics; NHIS-HEALS, National Health Insurance Service National Health Screening.

VESALIUS-CV to identify patients with chronic kidney disease (CKD) and was translated to CKD stage 3 or 4 in VESALIUS-REAL. The clinical trial's age criterion of 55–80 in women and 50–80 in men was streamlined to 50+ across both sexes to simplify the age criteria with limited impact on the cohort overall. Third, a number of clinical trial criteria could not be translated into real-world data and were therefore not used in the patient identification. These include coronary artery calcium (CAC) scores, ABI or lipoprotein(a) measurements. Despite these modifications to the VESALIUS-CV criteria, we are confident to capture a similar cohort of high-risk pre-event patients in the real world across a number of databases.

Table 3 Examples of translations of VESALIUS-CV to VESALIUS-REAL eligibility criteria

VESALIUS-CV eligibility criterion	VESALIUS-REAL eligibility criterion	Rationale for VESALIUS-REAL eligibility criterion
Translation of criterion with no changes		
≥65 years of age	Age ≥65 years	
MI or stroke before randomisation	Use native code* for diagnosis/procedure for MI or stroke	
Adjusted VESALIUS-CV criterion to VESALIUS-REAL		
Significant atherosclerotic CeVD meeting ≥1 of the following: (1) prior TIA with ≥50% carotid stenosis or (2) internal or external carotid artery stenosis of ≥70% or ≥ 2 of ≥ 50% stenoses or (3) prior internal or external carotid artery revascularisation	Use native code* for diagnosis/procedure of TIA, stenosis of carotid artery or cerebral artery syndrome	► Diagnosis/procedure codes for TIA, stenosis of carotid artery or cerebral artery syndrome available in most databases ► % stenosis not available in the database
Known FH	Use native code* for diagnosis/procedure of FH (definitive FH patient) or probable FH patient based on modified DLCN (score ≥6 points) when excluding DNA analysis or family history	Diagnosis codes for FH not available in most databases or likely underdiagnosed condition
eGFR 15 to <45 mL/min/1.73 m^2	Use native code* for diagnosis/procedure for CKD stage 3 or 4	eGFR not available in most databases
Translation of VESALIUS-CV criterion not possible for VESALIUS-REAL		
Lp(a) > 125 nmol/L (50 mg/dL)	Not included	Lp(a) not available in most databases
CAC score ≥300 in a patient without a coronary revascularisation	Not included	CAC score not available in most databases
ABI <0.85	Not included	ABI not available in the database

Online supplemental table 3 shows a full list of translations of clinical trial to observational study eligibility criteria.
*Native codes included diagnosis or procedure codes in ICD-10 or equivalent used in local coding vocabulary.
ABI, ankle-brachial index; CAC, coronary artery calcium; CeVD, cerebrovascular disease; CKD, chronic kidney disease; CV, cardiovascular; DLCN, Dutch Lipid Clinic Network; eGFR, estimated glomerular filtration rate; FH, familial hypercholesterolemia; ICD-10, International Classification of Diseases Tenth Revision; Lp(a), lipoprotein (a); MI, myocardial infarction; TIA, transient ischaemic attack; VESALIUS-CV, Effect of Evolocumab in Patients at High Cardiovascular Risk Without Prior Myocardial Infarction or Stroke; VESALIUS-REAL, Global Burden Study of High-Risk (VESALIUS-CV like) Population in the REAL-World.

Patients will be included in the study if they meet the following inclusion criteria (online supplemental figure 1):

1. Evidence of one of the four medical conditions (coronary artery disease [CAD], CeVD, AD or diabetes mellitus with microvascular complications or chronic insulin use) and
2. At least one elevated lipid measurement (LDL-C ≥90 mg/dL [≥2.3 mmol/L], non–high-density lipoprotein cholesterol [non–HDL-C] ≥ 120 mg/dL [≥3.1 mmol/L], apolipoprotein B [apoB] ≥80 mg/dL [≥1.56 µmol/L]) and
3. At least one of the following high-risk criteria:
   a. Evidence of another medical condition such as CAD, CeVD, PAD and diabetes with microvascular complications or chronic insulin use
   b. The qualifying limit measurement is LDL-C ≥130 mg/dL (≥3.36 mmol/L), non–HDL-C ≥160 mg/dL (>4.14 mmol/L) and apoB ≥120 mg/dL (2.3 µmol/L)
   c. Age ≥65 years
   d. Familial hypercholesterolaemia
   e. Ever tobacco use
   f. CKD stage 3 or 4
   g. Premature menopause

For objectives 2–5, the index date of each patient will be defined as the earliest point in time when they meet the inclusion criteria within the site-specific study period. Patients will be excluded if they have evidence of MI, stroke or end-stage renal disease prior to their index date; are <50 years of age at the index date; have no information available regarding their sex; or have <2 years of observability in the database before the index date. For objective 1, patients' eligibility will be assessed on a year-by-year basis.

Baseline and follow-up period

The baseline period includes all medical history regardless of duration up to and including the index date. No follow-up is required for objectives 1 and 2. For objectives 3–5, each patient will be followed from the day following the index date until death, loss to follow-up or the end of data available in the database, whichever occurs first. For objective 4, follow-up will additionally end at a patient's first CV event defined as an endpoint in this study.

Endpoints, outcome and covariate assessment

Table 4 describes the endpoints for study objectives. Incidence of CV events following the index date will be assessed using medical codes for relevant CV events as well as death records containing date or month of death and, where available, cause of death. Lipid measurements will be assessed through laboratory data. Lipid-lowering medications will be identified through prescription/dispensing records. Global code lists using International Classification of Diseases, Tenth Revision (ICD-10), and WHO Anatomical Therapeutic Chemical (ATC) Classification System codes have been developed to define conditions and prescriptions used in this study (see online supplemental table 4 for codes). These code lists will be adapted by the owners of each database to the local coding vocabularies using local clinical expertise.

Statistical analysis

Data will be summarised descriptively. Statistical Analysis System (SAS) v.9.4 and Microsoft Excel will be used for data analyses and management. Continuous data will be summarised using mean, SD, median, IQR and range. Categorical variables will be presented as the number and proportion in each category. Rates will be presented using 95% CIs generated using Poisson methodology. For time-to-event endpoints, Kaplan-Meier (KM) curves and estimates at defined time points will be generated. Since the amount of missing data is expected to be small and its impact on the reliability of the results minimal, proportions of missing data will be reported, where relevant, and no data will be imputed. The objectives to evaluate are listed below.

Objective 1: prevalence of high-risk pre–CV-event individuals

For databases that provide patient-level population-wide data, the population at risk will be defined as people who are aged ≥50 years, are enrolled/active in the database and have ≥2 years of prior records in the given calendar year. Where population-wide data are not available in the database, published data will be used. Prevalence (expressed per 100 persons) per calendar year of VESALIUS-CV–like individuals will be calculated as the sum

Table 4 Study endpoints

Objectives	Endpoints
Primary*	
Objective 3: describe care pathways of VESALIUS-CV–like individuals	► Prescription of relevant lipid-lowering treatments following index date ► Relevant laboratory measures following index date
Objective 4: estimate the incidence rates of CV events among VESALIUS-CV–like individuals	Incident episodes of ► MI ► Stroke ► Ischaemia-driven arterial revascularisation† ► CHD death (where available) ► CV death (where available) ► Death from all causes ► Composite of MI, stroke, ischaemia-driven arterial revascularisation or all-cause death ► Composite of MI, stroke, ischaemia-driven arterial revascularisation, CV or CHD death (where available) ► Composite of non-fatal MI, non-fatal stroke or CV death (where available) ► Composite of non-fatal/fatal MI or non-fatal/fatal stroke ► Composite of MI, stroke, all-cause death ► Composite of MI, stroke, CV or CHD death (where available)
Exploratory	
Objective 5: estimate the healthcare costs (in US data only) of the VESALIUS-CV–like population in real world	► Baseline CV-related costs between the index date and first event ► Acute and long-term CV-related healthcare costs following MI, stroke or ischaemia-driven arterial revascularisation

*Objectives 1 and 2 describe prevalence and patient characteristics.
†Ischaemia-driven arterial revascularisation defined as any of the following: CABG/PCI, revascularisation procedures for PAD, revascularisation procedures for carotid artery stenosis, abdominal aorta bypass, abdominal thoracic bypass, carotid endarterectomy, thrombolysis procedures for arteries, on thrombolytic medication, renal artery revascularisation, mesenteric artery revascularisation, splenic artery revascularisation and hepatic artery revascularisation.
CABG, coronary artery bypass grafting; CHD, coronary heart disease; CV, cardiovascular; MI, myocardial infarction; PAD, peripheral artery disease; PCI, percutaneous coronary intervention; VESALIUS-CV, Effect of Evolocumab in Patients at High Cardiovascular Risk Without Prior Myocardial Infarction or Stroke.

of these individuals in the calendar year divided by the population at risk in the same calendar year. A linear regression model will be used to test for temporal trends in annual prevalence in each database, assuming a linear trend for prevalence throughout the study period. Annual prevalence as a dependent variable and calendar year as a predictor variable will be fitted into the model. A two-tailed $p<0.05$ will be considered statistically significant.

Objective 2: characteristics of high-risk pre–CV-event individuals

Descriptive statistics will be used to describe the baseline characteristics of patients in each database.

Objective 3: care pathways of high-risk pre–CV-event individuals

The description of LLT after the index date will include (1) number and proportions of patients receiving at least one LLT prescription within specified time intervals following the index date (1 and 2 years); (2) number and proportions of patients having at least one lipid test recorded within specific time intervals (1 and 2 years); (3) number and proportions of patients with elevated lipid records (defined as LDL-C >70 mg/dL [1.8 mmol/L]) among those with at least one lipid test within the specific time intervals (1 and 2 years); (4) number and proportions of patients who receive treatment intensification at specific time intervals (3 and 12 months) and (5) for patients with a recorded lipid test following the index date (1 and 2 years), proportions of patients receiving treatment intensification at specific time intervals (1 and 2 years) following the date of the first elevated lipid record (defined as LDL >70 mg/dL [1.8 mmol/L]) will be described. The LLT agents considered in this study are statins, PCSK9 inhibitors, bempedoic acid, ezetimibe and other LLTs, such as bile acid sequestrants, niacin, fibrates, lomitapide and mipomersen.

Objective 4: incidence of CV events

The population at risk will be VESALIUS-CV–like individuals. Number of incident events (per 1000 person-years) of each specified CV event, mortality and composite endpoints will be calculated as the sum of each first incident event divided by the person-years of the population risk multiplied by 1000. Additionally, incidence rates will be computed 3 and 5 years from the time of the index date using the KM methodology. Results will be stratified by sex and age (in 5-year age bands, including: 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84 and 85+ years). To allow for cross-geographical comparisons, estimates will be age and sex standardised[12] using 2020 United Nations world population estimates[13] as standard. Endpoints and composites that require cause of death information will be estimated only in databases where such data are available.

Objective 5: healthcare costs

Costs will be summarised per person and year, for three different time periods: (1) baseline (defined as time from the index date to initial event), (2) acute (defined as the first year following the initial event) and (3) long term (defined as the second year following the initial event). Costs will be broken down into those incurred through billing records and those incurred through pharmacy.

Additional analysis

Sensitivity analyses will be conducted to evaluate the robustness of the analysis. To describe changes over time, analysis for selected objectives will be conducted for different time periods including 2010–2016 before versus 2017–2020 after evolocumab approval. For objective 3, a subgroup analysis will be performed for patients who have their LDL-C measure as the last event to define their index date. Additional subgroup analysis may be conducted on cohorts of interest, including but not limited to diabetes, a key subgroup of VESALIUS-CV–like patients.

Sample size

This study is descriptive, and no formal sample size calculations will be performed. Each database is estimated to provide a sufficiently large population to identify a pool of VESALIUS-CV–like individuals to be followed until they experience relevant incident CV events (table 2).

Limitations

Databases in this study were built for clinical administrative or reimbursement purposes, rather than for epidemiological research. The databases represent a variety of data sources, healthcare settings and coding practices, each of which will have different features and limitations. The strengths and limitations of different types of databases have been discussed elsewhere.[14 15]

Measurement errors/misclassification

While the datasets in this study have been carefully selected, individual datasets capture only part of all treatments provided in a given geography, for instance, by only capturing publicly or privately reimbursed treatments. However, it is expected that this will have only limited impact on the validity of the results as each dataset captures at least the main care setting.

Diagnosis and recording practice vary between individual countries and are captured to different levels of granularity. It is therefore possible that patients are misclassified; however, the phenotypes have been created in a way that is inclusive across the data sources in this study.

The study will use prescription data to assess treatment, which is only a proxy for the dispersion or the patient taking their medication. Actual treatment with certain medications may therefore be overestimated.

Cause of death is used in objective 4. Depending on the robustness of data collection for CV death in the respective database, there is a risk of misclassification. This misclassification is limited as studies from Sweden[16] and the United Kingdom[17 18] show a high level of concordance between medical records and the cause of death register in the respective countries.

Diagnosis and coding practices are likely to vary between countries, making cross-geographical comparisons difficult. To reduce the impact of these differences, code lists for relevant diagnosis have been designed to be inclusive to ensure diagnoses are captured in a comprehensive way.

Potential study biases

All patients who fulfil the eligibility criteria in each database will be included. Databases in this study have been selected for their representativeness of the population, reducing the impact of selection bias. A potential source of selection bias can be seen in the inclusion of laboratory measurements (including lipid measurements) in the inclusion criteria. While the databases in this study have been selected to provide good laboratory data coverage, not all patients will have valid lipid measurements and only a subset of patients with elevated lipid measurements will be included in the study. Of note, guidelines and care pathways may have changed over the duration of the study.

A particular strength of this study includes the use of a global protocol applied across all databases as well as a design of an ACDM that allows for common statistical programming to be applied across all databases. While emulating clinical trial eligibility criteria to real-world data can be challenging, we have applied an approach to ensure real-world results are as comparable to the clinical trial population as possible.

Patient and public involvement

The study will retrospectively analyse secondary data. Patients are deidentified, so direct patient involvement will not be possible. However, several researchers involved in this study and in a study-specific advisory board routinely consult with patients in the design, development and reporting of research at a national level. Patients may be involved in presentations and dissemination of the results at a national level.

ETHICS AND DISSEMINATION

Approvals of the study protocol were obtained from relevant local ethics and regulatory frameworks for each participating database. The institutions granting approval in each country and the status and approval numbers for each database are listed in online supplemental table 5. All data contributing to this study are taken from existing anonymised patient records. Only aggregated data will be shared. The results of the study will be submitted to peer-reviewed scientific publications and presented at scientific conferences. Authorship of any publications resulting from this study will be determined on the basis of the International Committee of Medical Journal Editors Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

Author affiliations
[1]Center for Observational Research, Amgen Ltd, Uxbridge, UK
[2]Center for Observational Research, Amgen Inc, Thousand Oaks, California, USA
[3]Global Development, Amgen Inc, Thousand Oaks, California, USA
[4]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China
[5]Laboratory of Data Discovery for Health (D24H), Hong Kong Science and Technology Park, Sha Tin, Hong Kong SAR, China
[6]School of Nursing, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China
[7]Advanced Data Analytics for Medical Science (ADAMS) Limited, Hong Kong SAR, China
[8]Institute of Clinical Pharmacy and Pharmaceutical Sciences, National Cheng Kung University, Taiwan, Taiwan
[9]Population Health Data Center, National Cheng Kung University, Tainan, Taiwan
[10]Sence Research, Uppsala, Sweden
[11]School of Pharmacy, Sungkyunkwan University—Suwon Campus, Suwon, Gyeonggi-do, Republic of Korea
[12]PHARMO Institute, Utrecht, The Netherlands
[13]Aston Pharmacy School, Aston University, Birmingham, UK

Contributors AO: conceptualisation, methodology, writing—original draft preparation. QC, AYLC, ICKW: conceptualisation, methodology, writing—reviewing and editing. NND, JO'K: conceptualisation, methodology, writing—reviewing and editing, supervision. MD, EDS: software, methodology, writing—reviewing and editing. GPDSL, SA, EMH: writing—reviewing and editing. CSLC: conceptualisation, methodology, data curation, writing—reviewing and editing. EC-CL, TC, J-YS: data curation, writing—reviewing and editing. AO is the guarantor of the manuscript.

Funding This work was supported by Amgen Inc. Medical writing support was provided by Martha Mutomba, PhD, and funded by Amgen Inc. Amgen Inc was involved in the design of the study reported in this manuscript. Thomas Cars is a co-owner of Sence Research that received funding from Amgen to fund this work. Edith Heintjes is an employee of the PHARMO Institute for Drug Outcomes Research that received funding from Amgen to fund this work.

Competing interests AO, QC, MD and JO'K are employees of Amgen Ltd and hold stock in Amgen Inc. NND, ES, GPDSL and SA are employees of and hold stock in Amgen Inc. AYLC and EC-CL have no conflicts to disclose. CSLC received research/grant support (paid to institution) from Food and Health Bureau of the Hong Kong Government, Hong Kong Research Grant Council, Hong Kong Innovation and Technology Commission, Pfizer, IQVIA, MSD, and Amgen; received consulting fees from PrimeVigilance; and was a non-executive director of Advance Data Analytics for Medical Science (ADAMS) Limited (HK). TC is a co-owner and holds stock of Sence Research, which is an independent company in epidemiology and biostatistics with both healthcare authorities and private companies as clients and has received payments from different clients (no payments were made directly to TC). Sence Research received funding from Amgen to fund this work. J-YS received grants from the Ministry of Food and Drug Safety, Ministry of Health and Welfare, National Research Foundation of Korea, and Government-wide R&D Fund for Infectious Disease Research and from Pfizer, LG Chemical, and Union Chimique Belge. EMH was an employee of the PHARMO Institute for Drug Outcomes Research. This independent research institute performs financially supported studies for government and related healthcare authorities and several pharmaceutical companies including Amgen Inc.ICKW received research/grant support (paid to institution) from National Health and Medical Research Council in Australia, Bristol Myers Squibb, Pfizer, Janssen, Bayer, GlaxoSmithKline, Novartis, The Hong Kong Research Grants Council, Health Bureau of the Government of the Hong Kong Special Administrative Region, National Institute for Health Research in England and the European Commission; received consulting fees from the WHO, IQVIA and Jacobson Medical in Hong Kong (non-executive director); received payment for expert testimony (payment made to institution) from the Hong Kong Appeal Court; is a member of the Expert Committee on Clinical Events Assessment Following COVID-19 Immunization, Department of Health, The Government of the Hong Kong Special Administrative Region, Hong Kong Special Administrative Region, China; is the current founder and director of Advance Data Analytics for Medical Science (ADAM) Limited (HK), Asia Medicine Regulatory Affairs (AMERA) Services Limited and OCUS Innovation Limited (HK, Ireland and UK); and is a former director of Therakind Limited (UK).

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been

ORCID iDs
Andreas Ochs http://orcid.org/0009-0000-3434-2090
Queenie Chan http://orcid.org/0000-0001-9278-230X
Michael Duxbury http://orcid.org/0009-0004-9403-7708
Erin Shannon http://orcid.org/0000-0003-2241-264X
James O'Kelly http://orcid.org/0000-0002-0718-9286
Gabriel Paiva da Silva Lima http://orcid.org/0000-0001-8073-719X
Celine SL Chui http://orcid.org/0000-0003-1513-8726
Edward Chia-Cheng Lai http://orcid.org/0000-0002-5852-7652
Thomas Cars http://orcid.org/0009-0008-0912-8226
Ju-Young (Judy) Shin http://orcid.org/0000-0003-1010-7525
Edith M Heintjes http://orcid.org/0000-0002-2617-3845

REFERENCES

1 World Health Organization. Cardiovascular diseases (CVDs) fact sheet. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [Accessed 22 Jun 2022].
2 Ference BA, Ginsberg HN, Graham I, *et al*. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2017;38:2459–72.
3 Magnussen C, Ojeda FM, Leong DP, *et al*. Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality. *N Engl J Med* 2023;389:1273–85.
4 Tian F, Chen L, Qian ZM, *et al*. Ranking age-specific modifiable risk factors for cardiovascular disease and mortality: evidence from a population-based longitudinal study. *EClinicalMedicine* 2023;64:102230.
5 Silverman MG, Ference BA, Im K, *et al*. Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-analysis. *JAMA* 2016;316:1289–97.
6 Cholesterol Treatment Trialists (CTT). The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *The Lancet* 2012;380:581–90.
7 Sabatine MS, Giugliano RP, Keech AC, *et al*. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* 2017;376:1713–22.
8 Bohula EA, Marston NA, Ruzza A, *et al*. Rationale and design of the effect of evolocumab in patients at high cardiovascular risk without prior myocardial infarction or stroke (VESALIUS-CV) trial. *Am Heart J* 2024;269:179–90.
9 Lai EC-C, Ryan P, Zhang Y, *et al*. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. *Clin Epidemiol* 2018;10:875–85.
10 Franklin JM, Pawar A, Martin D, *et al*. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clin Pharmacol Ther* 2020;107:817–26.
11 Nordestgaard BG, Chapman MJ, Humphries SE, *et al*. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur Heart J* 2013;34:3478–90.
12 Ahmad OB, Boschi-Pinto C, Lopez AD, *et al*. Age standardization of rates: a new WHO standard. GPE discussion paper series. World Health Organization; 2001.
13 United Nations. *World population prospects 2022: summary of results*.Available: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf [Accessed 22 Jun 2025].
14 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
15 Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol* 2021;21:234.
16 Eriksson A, Stenlund H, Ahlm K, *et al*. Accuracy of death certificates of cardiovascular disease in a community intervention in Sweden. *Scand J Public Health* 2013;41:883–9.
17 Wang X, Mobley AR, Tica O, *et al*. Systematic approach to outcome assessment from coded electronic healthcare records in the DaRe2THINK NHS-embedded randomized trial. *Eur Heart J Digit Health* 2022;3:426–36.
18 Gallagher AM, Dedman D, Padmanabhan S, *et al*. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf* 2019;28:563–9.