BMJ Open

Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies

Rohan Khera ,[1,2] Martijn J Schuemie ,[3,4] Yuan Lu ,[1,2] Anna Ostropolets ,[5] RuiJun Chen,[6] George Hripcsak,[5,7] Patrick B Ryan,[3,5] Harlan M Krumholz ,[1,2] Marc A Suchard [4,8,9,10]

To cite: Khera R, Schuemie MJ, Lu Y, *et al*. Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies. *BMJ Open* 2022;**12**:e057977. doi:10.1136/bmjopen-2021-057977

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057977).

Received 04 November 2021
Accepted 27 April 2022

For numbered affiliations see end of article.

Correspondence to
Marc A Suchard;
msuchard@ucla.edu

ABSTRACT

Introduction Therapeutic options for type 2 diabetes mellitus (T2DM) have expanded over the last decade with the emergence of cardioprotective novel agents, but without such data for older drugs, leaving a critical gap in our understanding of the relative effects of T2DM agents on cardiovascular risk.

Methods and analysis The large-scale evidence generations across a network of databases for T2DM (LEGEND-T2DM) initiative is a series of systematic, large-scale, multinational, real-world comparative cardiovascular effectiveness and safety studies of all four major second-line anti-hyperglycaemic agents, including sodium–glucose co-transporter-2 inhibitor, glucagon-like peptide-1 receptor agonist, dipeptidyl peptidase-4 inhibitor and sulfonylureas. LEGEND-T2DM will leverage the Observational Health Data Sciences and Informatics (OHDSI) community that provides access to a global network of administrative claims and electronic health record data sources, representing 190 million patients in the USA and about 50 million internationally. LEGEND-T2DM will identify all adult, patients with T2DM who newly initiate a traditionally second-line T2DM agent. Using an active comparator, new-user cohort design, LEGEND-T2DM will execute all pairwise class-versus-class and drug-versus-drug comparisons in each data source, producing extensive study diagnostics that assess reliability and generalisability through cohort balance and equipoise to examine the relative risk of cardiovascular and safety outcomes. The primary cardiovascular outcomes include a composite of major adverse cardiovascular events and a series of safety outcomes. The study will pursue data-driven, large-scale propensity adjustment for measured confounding, a large set of negative control outcome experiments to address unmeasured and systematic bias.

Ethics and dissemination The study ensures data safety through a federated analytic approach and follows research best practices, including prespecification and

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The proposal seeks to use health information encompassing millions of patients with type 2 diabetes mellitus (T2DM) in the multinational Observational Health Data Science and Informatics (OHDSI) community to determine real-world comparative effectiveness and safety of traditionally second-line T2DM agents.

⇒ The proposed set of studies will be comprehensive, with a systematic pairwise comparisons of all sodium–glucose co-transporter-2 inhibitor, glucagon-like peptide-1 receptor agonist, dipeptidyl peptidase-4 inhibitor and sulfonylurea agents at the drug, class and population subgroup level.

⇒ The studies will focus on a broad set of outcomes, including comprehensive measures of adverse cardiovascular events as well as secondary effectiveness and safety outcomes.

⇒ The studies use robust methods an observational, active-comparator, new-user cohort design with a systematic framework to address residual confounding, publication bias and p-hacking using data-driven, large-scale propensity adjustment for measured confounding, a large set of negative control outcome experiments to address unmeasured and systematic bias, prespecification and full disclosure of hypotheses tested and their results. These approaches capitalise on mature OHDSI open-source resources and a large body of clinical and quantitative research that the LEGEND-T2DM investigators originated and continue to drive.

⇒ The study will focus on drug effectiveness rather than safety without the ability to systematically track the adherence to individual agents across cohorts.

full disclosure of results. LEGEND-T2DM is dedicated to open science and transparency and will publicly share all analytic code from reproducible cohort definitions through turn-key software, enabling other research groups to leverage our methods, data and results to verify and extend our findings.

RATIONALE AND BACKGROUND

The landscape of therapeutic options for type 2 diabetes mellitus (T2DM) has been dramatically transformed over the last decade.[1] The emergence of drugs targeting the sodium–glucose co-transporter-2 (SGLT2) and the glucagon-like peptide-1 (GLP1) receptor has expanded the role of T2DM agents from lowering blood glucose to directly reducing cardiovascular risk.[2] A series of large randomised clinical trials designed to evaluate the cardiovascular safety of SGLT2 inhibitors (SGLT2Is) and GLP1 receptor agonists (GLP1RAs) found that use of many of these agents led to a reduction in major adverse cardiovascular events, including myocardial infarction, hospitalisation for heart failure and cardiovascular mortality.[3–6] However, other T2DM drugs widely used before the introduction of these novel agents, such as sulfonylureas, did not undergo similarly comprehensive trials to evaluate their cardiovascular efficacy or safety. Moreover, direct comparisons of newer agents with dipeptidyl peptidase-4 (DPP4) inhibitors (DPP4Is), with neutral effects on major cardiovascular outcomes,[7–10] have not been conducted. Nevertheless, DPP4Is and sulfonylureas continue to be used in clinical practice and are recommended as second-line T2DM agents in national clinical practice guidelines.

Several challenges remain in formulating T2DM treatment recommendations based on existing evidence.[11] First, trials of novel agents did not pursue head-to-head comparisons to older agents and were instead designed as additive treatments on the background of commonly used T2DM agents. Therefore, the relative cardiovascular efficacy and safety of novel agents compared with older agents is not known, and indirect estimates have relied on summary-level data restricted to common comparators[12–14] and are less reliable.[15 16] Second, trials of novel agents have tested individual drugs against placebo but have not directly compared SGLT2Is with GLP1RAs in reducing adverse cardiovascular event risk. Moreover, there is no evidence to guide the use of individual drugs within each class and across different drug classes, particularly among patients at lower cardiovascular risk than recruited in clinical trials. Third, randomised trials focused on cardiovascular efficacy and safety but were not powered to adequately assess the safety of these agents across a spectrum of non-cardiovascular outcomes. Finally, restricted enrolment across regions, and subgroups of age, sex and race further limits the efficacy and safety assessment that may guide individual patients' treatment.

Evidence gaps from these trials also pose a challenge in designing treatment algorithms, which rely on comparative effectiveness and safety of drugs. Perhaps, as a result, there is large variation in clinical practice guidelines and in clinical practice with regard to these medications, with many patients initiated on the newer therapies and many others treated with older regimens.[17–21] Among the second-line options, there is much variation with respect to the order of drugs used. This lack of consensus about the best approach provides an opportunity for systematic, large-scale observational studies.

STUDY OBJECTIVES

To inform critical decisions facing patients with diabetes, their caregivers, clinicians, policy-makers and healthcare system leaders, we have launched the large-scale evidence generation and evaluation across a network of databases for T2DM (LEGEND-T2DM) initiative to execute a series of comprehensive observational studies to compare cardiovascular outcome rates and safety of second-line T2DM glucose-lowering agents. Specifically, these studies aim:

1. To determine, through systematic evaluation, the comparative effectiveness of traditionally second-line T2DM agents, SGLT2Is and GLP1RAs, with each other and with DPP4Is and sulfonylureas, for cardiovascular outcomes.
2. To determine, through systematic evaluation, the comparative safety of traditionally second-line T2DM agents among patients with T2DM.
3. To assess heterogeneity in effectiveness and safety of traditionally second-line T2DM agents among key patient subgroups: using stratified patient cohorts, we will quantify differential effectiveness and safety across subgroups of patients based on age, sex, race, renal impairment and baseline cardiovascular risk.

RESEARCH METHODS

LEGEND-T2DM will execute three systematic, large-scale observational studies of second-line T2DM agents to estimate the relative risks of cardiovascular effectiveness and safety outcomes.

1. The Class-versus-class study will provide all pairwise comparisons between the four major T2DM agent classes to evaluate their comparative effects on cardiovascular risk (Objective 1) and patient-centred safety outcomes (Objective 2).
2. The drug-versus-drug study will furnish head-to-head pairwise comparisons between individual agents within and across classes (both Objectives 1 and 2).
3. The heterogeneity study will refine these comparisons for patients with T2DM for important subgroups (Objective 3). In contrast to a single comparison approach, LEGEND-T2DM will provide a comprehensive view of the findings and their consistency across populations, drugs and outcomes. We will model each study on our successful collaborative research evaluating the comparative effectiveness of antihypertensives recently published in *The Lancet*.[22]

Table 1 list the four major T2DM agent classes and the individual agents licensed in the USA within each class.

Table 1 T2DM drug classes and individual agents within each class

DPP4 inhibitors	GLP1 receptor antagonists	SGLT2 inhibitors	Sulfonylureas
Alogliptin	Albiglutide	Canagliflozin	Chlorpropamide
Linagliptin	Dulaglutide	Dapagliflozin	Glimepiride
Saxagliptin	Exenatide	Empagliflozin	Glipizide
Sitagliptin	Liraglutide	Ertugliflozin	Gliquidone
Vildagliptin	Lixisenatide		Glyburide
	Semaglutide		Tolazamide
			Tolbutamide

DPP4, dipeptidyl peptidase-4; GLP1, glucagon-like peptide-1; SGLT2, sodium–co-transporter-2; T2DM, type 2 diabetes mellitus.

We will examine all $\binom{4}{2} = 6$ class-wise comparisons and all $\binom{\frac{5+6+4+7}{2}}{2} = 231$ ingredient-wise comparisons. For each comparison, we are interested in the relative risk of each of the cardiovascular and safety outcomes described in the Outcomes section.

Study design

For each study, we will employ an active comparator, new-user cohort design.[23–25] New-user cohort design is advocated as the primary design to be considered for comparative effectiveness and drug safety.[26–28] By identifying patients who start a new treatment course and using therapy initiation as the start of follow-up, the new-user design models a randomised controlled trial (RCT), where treatment commences at the index study visit. Exploiting such an index date allows a clear separation of baseline patient characteristics that occur prior to index date and are usable as covariates in the analysis without concern of inadvertently introducing mediator variables that arise between exposure and outcome.[29] Excluding prevalent users as those without a sufficient washout period prior to first exposure occurrence further reduces bias due to balancing mediators on the causal pathway, time-varying hazards and depletion of susceptibles.[28 30] Our systematic framework across studies further will address residual confounding, publication bias and p-hacking using data-driven, large-scale propensity adjustment for measured confounding,[31] a large set of negative control outcome experiments to address unmeasured and systematic bias[32–34] and full disclosure of hypotheses

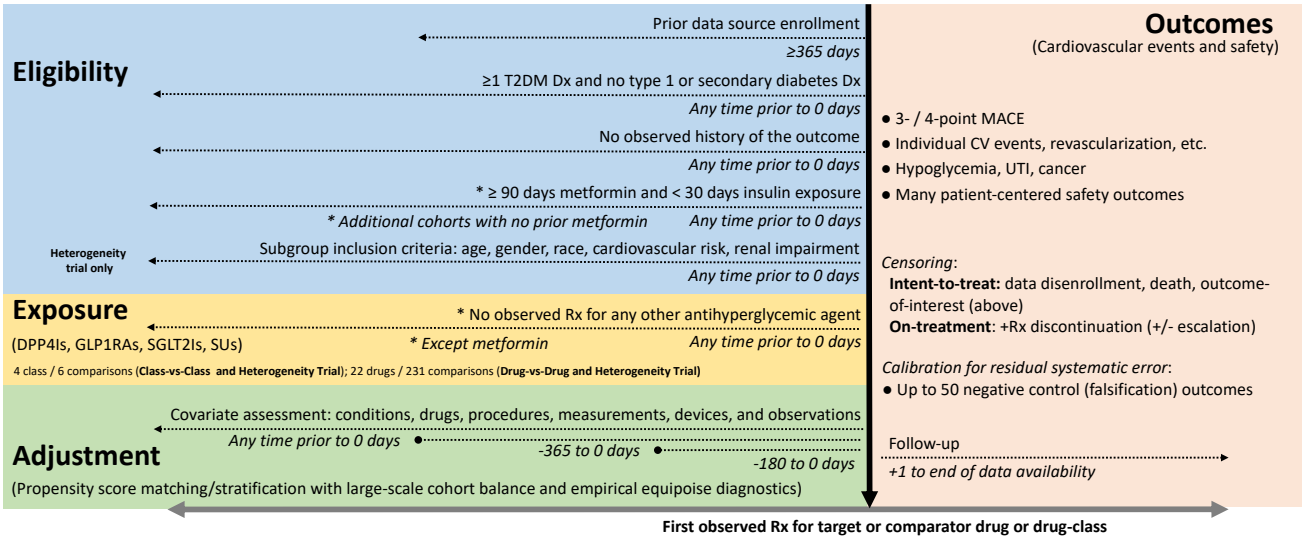

Figure 1 Schematic of LEGEND-T2DM new-user cohort design for the class-versus-class, drug-versus-drug and heterogeneity studies. DPP4Is, dipeptidyl peptidase-4 inhibitors; GLP1RAs, glucagon-like peptide-1 receptor agonists; LEGEND-T2DM, large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus; MACE, major adverse cardiovascular events; SGLT2Is, sodium–glucose co-transporter-2 inhibitors; CV, cardiovasuclar; UTI, urinary track infection

tested.[35] Figure 1 illustrates our design for all studies that the following sections describe in more detail.

Data sources

We will execute LEGEND-T2DM as a series of OHDSI network studies. All data partners within OHDSI are encouraged to participate voluntarily and can do so conveniently, because of the community's shared Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and OHDSI tool stack. Many OHDSI community data partners have already committed to participate and we will recruit further data partners through OHDSI's standard recruitment process, which includes protocol publication on OHDSI's GitHub, an announcement in OHDSI's research forum, presentation at the weekly OHDSI all-hands-on meeting and direct requests to data holders.

Table 2 lists the 13 already committed data sources for LEGEND-T2DM; these sources encompass a large variety of practice types and populations. For each data source, we report a brief description and size of the population it represents and its patient capture process and start date. While the earliest patient capture begins in 1989 (Columbia University Irving Medical Center, CUIMC), the vast majority come from the mid-2000s to today, providing almost two decades of T2DM treatment coverage. US populations include those commercially and publicly insured, enriched for older individuals (MDCR, VA), lower socioeconomic status (MDCD) and racially diverse (VA >20% Black or African American, CUIMC 8%). The US data sources may capture the same patients across multiple sources. Different views of the same patients are an advantage in capturing the diversity of real-world health events that patients experience. Across Commercial Claims and Encounters (CCAE; commercially insured), MCDR (Medicare) and MCDC (Medicaid), we expect little overlap in terms of the same observations recorded at the same time for a patient; patients can flow between sources (eg, a CCAE patient who retires can opt-in to become an MDCR patient), but the enrolment time periods stand distinct. On the other hand, Optum, PanTher, OpenClaims, CUIMC and Yale New Haven Health System may overlap in time with the other US data sources. While it remains against licensing agreements to attempt to link patients between most data sources, Optum reports <20% overlap between their claims and electronic health record (EHR) data sources that is reassuringly small. All data sources will receive institutional review board approval or exemption for their participation before executing LEGEND-T2DM.

Study population

We will include all subjects in a data source who meet inclusion criteria for one or more traditionally second-line T2DM agent exposure cohorts. Broadly, these cohorts will consist of patients with T2DM either with or without prior metformin monotherapy who initiate treatment with one of the 22 drug ingredients that comprise the DPP4I, GLP1RA, SGLT2I and sulfonylurea drug classes (table 1). We do not consider thiazolidinediones, given their known association with a risk of heart failure and bladder cancer.[36 37] We describe specific definitions for exposure cohorts for each study in the following sections.

Exposure comparators

Class-versus-class study comparisons

The class-versus-class study will construct four exposure cohorts for new users of any drug ingredient within the four traditionally second-line drug classes in table 1. Cohort entry (index date) for each patient is their first observed exposure to any drug ingredient for the four second-line drug classes. Consistent with an idealised target trial for T2DM therapy and cardiovascular risk,[38 39] inclusion criteria for patients based on the index date will include:

- T2DM diagnosis and no type 1 or secondary diabetes mellitus diagnosis before the index date;
- At least 1 year of observation time before the index date (to improve new-user sensitivity).
- No prior drug exposure to a comparator second-line or other antihyperglycaemic agent (ie, thiazolidinediones, acarbose, acetohexamide, bromocriptine, glibornuride, miglitol and nateglinide) or >30 days insulin exposure before index date.

We will construct and compare separately cohorts patients either with:

- At least 3 months of metformin use before the index date.
- No prior metformin use before the index date.

or

- No prior metformin use before the index date.

In the first case, 3 months of metformin is consistent with ADA guidelines.[40] In the second case, we are interested in relative effectiveness and safety of these traditionally second-line agents in patients who initiate their treatments without first using metformin. We purposefully do not automatically exclude or restrict to patients with a history of myocardial infarction, stroke or other major cardiovascular events, which will allow us to report relative effectiveness and safety for individuals with both low or moderate and high cardiovascular risk. Likewise, we do not automatically exclude or restrict to individuals with severe renal impairment.[41] We will use cohort diagnostics, such as achieving covariate balance and clinical empirical equipoise between exposure cohorts (see the Sample size and study power section) and stakeholder input to guide the possible need to exclude other prior diagnoses, such as congestive heart failure, pancreatitis or cancer.[41]

Online supplemental appendix A.1 reports the complete OHDSI ATLAS cohort description for new users of DDP4 inhibitors with prior metformin use. This description lists complete specification of cohort entry events, additional inclusion criteria, cohort exit events and all associated standard OMOP CDM concept code sets used in the definition. We generate programmatically

Table 2 Committed LEGEND-T2DM data sources and the populations they cover

Data source	Population	Patients (million)	History	Data capture process and short description
IBM MarketScan CCAE	Commercially insured, <65 years	142M	2000–today	Adjudicated health insurance claims (eg, inpatient, outpatient and outpatient pharmacy) from large employers and health plans who provide private healthcare coverage to employees, their spouses and dependents
IBM MarketScan Medicare Supplemental Database (MDCR)	Commercially insured, 65+ years	10M	2000–today	Adjudicated health insurance claims of retirees with primary or Medicare supplemental coverage through privately insured fee-for-service, point-of-service or capitated health plans
IBM MarketScan Multi-State Medicaid Database (MDCD)	Medicaid enrollees, racially diverse	26M	2006–today	Adjudicated health insurance claims for Medicaid enrollees from multiple states and includes hospital discharge diagnoses, outpatient diagnoses and procedures, and outpatient pharmacy claims
IQVIA Open Claims (IOC)	General	160M	2010–today	Pre-adjudicated claims at the anonymised patient-level collected from office-based physicians and specialists via office management software and clearinghouse switch sources for the purpose of reimbursement
JMDC	Japan, general	5.5M	2005–today	Data from 60 society-managed health insurance plans covering workers aged 18–65 years and their dependents
Korea NHIS	2% random sample of South Korea	1M	2002–today	National administrative claims database covering the South Korean population
Optum Clinformatics Data Mart (Optum)	Commercially or Medicare insured	85M	2000–today	Inpatient and outpatient healthcare insurance claims
CUIMC	Academic medical centre patients, racially diverse	6M	1989–today	General practice, specialists and inpatient hospital services from the New York-Presbyterian hospital and affiliated academic physician practices in New York
Department of Veterans Affairs (VA)	Veterans, older, racially diverse	12M	2000–today	National VA healthcare system, the largest integrated provider of medical services in the USA, provided at 170 VA medical centres and 1063 outpatient sites
Information System for Research in Primary Care (SIDIAP)	80% of all Catalonia (Spain)	7.7M	2006–today	Primary care partially linked to inpatient data with pharmacy dispensations and primary care laboratories. Healthcare is universal and taxpayer funded in the region, and PCPs are gatekeeps for all care and responsible for repeat prescriptions
IQVIA Disease Analyzer	Germany, general	37M	1992–today	Collection from patient management software used by general practitioners and selected specialists to document patients' medical records within their office-based practice during a visit
OptumEHR	The USA, general	93M	2006–today	Clinical information, prescriptions, lab results, vital signs, body measurements, diagnoses and procedures derived from clinical notes using natural language processing
YNHHS	Academic medical centre patients	2M	2013–today	General practice, specialists and inpatient hospital services from the YNHHS in Connecticut

CCAE, Commercial Claims and Encounters; CUIMC, Columbia University Irving Medical Center; IOC, IQVIA Open Claims; JMDC, Japan Medical Data Center; LEGEND-T2DM, large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus; NHIS, National Health Insurance Service; OptumEHR, Optum electronic health records; PCP, primary care physician; YNHHS, Yale New Haven Health System.

equivalent cohort definitions for new others of each drug class with and without prior metformin use. ATLAS then automatically translates these definitions into network-deployable SQL source code. Online supplemental appendix A.2 lists the inclusion criteria modifier for no prior metformin use.

Of note, the inclusion criteria do not directly incorporate quantitative measures of poor glycaemic control, such as one or more elevated serum hemoglobin A1c (HbA1c) measurements; such laboratory values are irregularly captured in large claims and even EHR data sources. Older ADA guidelines (but not since 2020 for patients with cardiovascular disease (CVD)[42]) advise escalating to a second-line agent only when glycaemic control is not met with metformin monotherapy, nicely mirroring our cohort design for our historical data. We will conduct sensitivity analyses involving available HbA1c measurements to demonstrate their balance between exposure cohorts (described later in the Sample size and study power section). In the unlikely event that balance is not met, we will consider an inclusion criterion of at least two HbA1c measurements ≥7% within 6 months before the index.[39] We will also conduct sensitivity analyses to assess prior insulin use exclusions, bearing in mind difficulties in assessing insulin use end-dates.

For each data source, we will then execute all $2 \times \binom{4}{2} = 6$ pairwise class comparisons for which the data source yields ≥1000 patients in each arm. Significantly fewer numbers of patients strongly suggest data source-specific differences in prescribing practices that may introduce residual bias and sufficient samples sizes are required to construct effective propensity score (PS) models.[43]

Drug-versus-drug study comparisons

The drug-versus-drug study will construct 2×22 exposure cohorts for new users of each drug ingredient in table 1. We will apply the same cohort definition, inclusion criteria and patient count minimum as described in the Class-versus-class study comparisons section.

For each data source, we will then execute all $2 \times \binom{22}{2} = 462$ pairwise drug comparisons. While we will publicly report studies results for all pairwise comparisons, we will focus primary clinical interpretation and scientific publishing to the $2 \times \binom{5}{2}$ (within DPP4Is) $+2 \times \binom{6}{2}$ (within GLP1RAs) $+2 \times \binom{4}{2}$ (within SGLT2Is) $+2 \times \binom{7}{2}$(within SUs) = 104 comparisons that pit drugs within the same class against each other, as well as across-class comparisons that stakeholders deem pertinent given their experiences.

Online supplemental appendix A.3 reports the complete OHDSI ATLAS cohort description for new users of aloglipitin with prior metformin use. Again, we programmatically construct all new-user drug-level cohort and automatically translate into SQL.

Heterogeneity study comparisons

The heterogeneity study will further stratify all 237 class-level and drug-level exposure cohorts in the Class-versus-class study comparisons section and the Drug-versus-drug study comparisons section by clinically important patient characteristics that modify cardiovascular risk or relative treatment heterogeneity to provide patient-focused treatment recommendations. These factors will include:

- • Age (18–44 years/45–64 years/≥65 years at the index date).
- Gender (women/men).
- Race (African American or black).
- Cardiovascular risk (low or moderate/high, defined by established CVD at the index date).
- Renal impairment (at the index date).

We will define patients at high cardiovascular risk as those who fulfil at index date an established CVD definition that has been previously developed and validated for risk stratification among new users of second-line T2DM agents.[44] Under this definition, established CVD means having at least one diagnosis code for a condition indicating CVD, such as atherosclerotic vascular disease, cerebrovascular disease, ischaemic heart disease or peripheral vascular disease, or having undergone at least one procedure indicating CVD, such as percutaneous coronary intervention, coronary artery bypass graft or revascularisation, any time on or prior to the exposure start. Likewise, we will define renal impairment through diagnosis codes for chronic kidney disease and end-stage renal disease, dialysis procedures and laboratory measurements of estimated glomerular filtration rate, serum creatinine and urine albumin.

Online supplemental appendix A.4 presents complete OHDSI ATLAS specifications for these subgroups, including all standard OMOP CDM concept codes defining cardiovascular risk and renal disease.

Validation

We will validate exposure cohorts and aggregate drug utilisation using comprehensive cohort characterisation tools against both claims and EHR data sources. Chief among these tools stands OHDSI's CohortDiagnostic package (github). For any cohort and data source mapped to OMOP CDM, this package systematically generates incidence new-user rates (stratified by age, gender and calendar year), cohort characteristics (all comorbidities, drug use, procedures and health utilisation) and the actual codes found in the data triggering the various rules in the cohort definitions. This can allow researchers and stakeholders to understand the heterogeneity of source coding for exposures and health outcomes as well as the impact of various inclusion criteria on overall cohort counts (details described in the Sample size and study power section).

Outcomes

Across all data sources and pairwise exposure cohorts, we will assess relative risks of 32 cardiovascular and

Table 3 LEGEND-T2DM study outcomes

Phenotype	Brief logical description	Prior development
3-point MACE	Condition record of acute myocardial infarction, haemorrhagic or ischaemic stroke or sudden cardiac death during an inpatient or ER visit	49–61
4-point MACE	3-point MACE+inpatient or ER visit (hospitalisation) with heart failure condition record	44 49–67
Acute myocardial infarction	Condition record of acute myocardial infarction during an inpatient or ER vist	49–54
Acute renal failure	Condition record of acute renal failure during an inpatient or ER visit	47 68–75
Glycaemic control	First haemoglobin A1c measurement with value ≤7%	76
Hospitalisation with heart failure	Inpatient or ER visit with heart failure condition record	44 62–67
Measured renal dysfunction	First creatinine measurement with value >3 mg/dL	75
Coronary revascularisation	Procedure record of percutaneous coronary intervention or coronary artery bypass grafting during an inpatient or ER visit	45
Stroke	Condition record of haemorrhagic or ischaemic stroke during an inpatient or ER visit	55–60
Sudden cardiac death	Condition record of sudden cardiac death during an inpatient or ER visit	52 61
Abnormal weight gain	Abnormal weight gain record of any type; successive records with >90-day gap are considered independent episodes; note, weight measurements not used	77
Abnormal weight loss	Abnormal weight loss record of any type; successive records with >90-day gap are considered independent episodes; note, weight measurements not used	78
Acute pancreatitis	Condition record of acute pancreatitis during an inpatient or ER visit	79–82
All-cause mortality	Death record of any type	52 83 84
Bladder cancer	Malignant tumour of urinary bladder condition record of any type; limited to earliest event per person	
Bone fracture	Bone fracture condition record of any type; successive records with >90-day gap are considered independent episodes	
Breast cancer	Malignant tumour of breast condition record of any type; limited to earliest event per person	
Diabetic ketoacidosis	Diabetic ketoacidosis condition record during an inpatient or ER visit	46 85
Diarrhoea	Diarrhoea condition record of any type; successive records with >30-day gap are considered independent episodes	86–88
GU infection	Condition record of any type of genital or urinary tract infection during an outpatient or ER vists	89
Hyperkalaemia	Condition record for hyperkalaemia or potassium measurements >5.6 mmol/L; successive records with >90-day gap are considered independent episodes	90–92
Hypoglycaemia	Hypoglycaemia condition record of any type; successive records with >90-day gap are considered independent episodes	93
Hypotension	Hypotension condition record of any type; successive records with >90-day gap are considered independent episodes	94
Joint pain	Joint pain condition record of any type; successive records with >90-day gap are considered independent episodes	
Lower extremity amputation	Procedure record of below knee lower extremity amputation during inpatient or outpatient visit	44 48
Nausea	Nausea condition record of any type; successive records with >30-day gap are considered independent episodes	95–97
Peripheral oedema	Oedema condition record of any type; successive records with >180-day gap are considered independent episodes	
Photosensitivity	Condition record of drug-induced photosensitivity during any type of visit	
Renal cancer	Primary malignant neoplasm of kidney condition record of any type; limited to earliest event per person	

Continued

Table 3 Continued

Phenotype	Brief logical description	Prior development
Thyroid tumour	Neoplasm of thyroid gland condition record of any type; limited to earliest event per person	
Venous thromboembolism	Venous thromboembolism condition record of any type; successive records with >180-day gap are considered independent episodes	98–101
Vomiting	Vomiting condition record of any type; successive records with >30-day gap are considered independent episodes	95–97

ER, emergence room; GU, genitourinary; LEGEND-T2DM, large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus; MACE, major adverse cardiovascular events.

patient-centred outcomes (table 3). Primary outcomes of interest are:

- Three-point major adverse cardiovascular events (MACE), including acute myocardial infarction, stroke and sudden cardiac death.
- Four-point MACE that additionally includes heart failure hospitalisation.

Secondary outcomes include:

- Individual MACE components.
- Acute renal failure.
- Coronary revascularisation.

In data sources with laboratory measurements, secondary outcomes further include:

- Glycaemic control.
- Measured renal dysfunction.

We will also study second-line T2DM drug side-effects and safety concerns highlighted in the 2018 ADA guidelines[40] and from RCTs, including:

- Abnormal weight change.
- Genitourinary infection.
- Various cancers.
- Hypoglycaemia.

We will employ the same level of systematic rigour in studying outcomes regardless of their primary or secondary label (online supplemental appendix B).

A majority of outcome definitions have been previously implemented and validated in our own work[22 44–48] based heavily on prior development by others (see references in table 3[44–101]). To assess across-source consistency and general clinical validity, we will characterise outcome incidence, stratified by age, sex and index year for each data source.

Analysis

Contemporary utilisation of drug classes and individual agents

For all cohorts in the three studies, we will describe overall utilisation as well as temporal trends in the use of each drug class and agents within the class. Furthermore, we will evaluate these trends in patient groups by age (18–44 years/45–64 years/≥65 years), gender, race and geographic regions. Since the emergence of novel medications in the management of T2DM in 2014, there has been a rapid expansion in both the number of drug classes and individual agents. These data will provide insight into the current patterns of use and possible disparities. These data are critical to guide the real-world application of treatment decision pathways for the treatment of patients with T2DM.

Specifically, we will calculate and validate aggregate drug utilisation using the OHDSI's CohortDiagnostic package against both claims and EHR data sources. The CohortDiagnostics package works in two steps: (1) generate the utilisation results and diagnostics against a data source and (2) explore the generated utilisation and diagnostics in a user-friendly graphical interface R-Shiny app. Through the interface, one can explore patient profiles of a random sample of subjects in a cohort. These diagnostics provide a consistent methodology to evaluate cohort definitions/phenotype algorithms across a variety of observational databases. This will enable researchers and stakeholders to become informed on the appropriateness of including specific data sources within analyses, exposing potential risks related to heterogeneity and variability in patient care delivery that, when not addressed in the design, could result in errors such as highly correlated covariates in PS matching of a target and a comparator cohort. Thus, the added value of this approach is twofold in terms of exposing data quality for a study question and ensuring face validity checks are performed on proposed covariates to be used for balancing PSs.

Relative risk of cardiovascular and patient-centred outcomes

For all three studies, we will execute a systematic process to estimate the relative risk of cardiovascular and patient-centred outcomes between new users of second-line T2DM agents. The process will adjust for measured confounding, control from further residual (unmeasured) bias and accommodate important design choices to best emulate the nearly impossible to execute, idealised RCT that our stakeholders envision across data source populations, comparators, outcomes and subgroups.

To adjust for potential measured confounding and improve the balance between cohorts, we will build large-scale PS models[102] for each pairwise comparison and data source using a consistent data-driven process through regularised regression.[31] This process engineers a large set of predefined baseline patient characteristics, including age, gender, race, index month/year and

other demographics and prior conditions, drug exposures, procedures, laboratory measurements and health service utilisation behaviours, to provide the most accurate prediction of treatment and balance patient cohorts across many characteristics. Construction of condition, drug, procedures and observations include occurrences within 365 days, 180 days and 30 days prior to index date and are aggregated at several SNOMED (conditions) and ingredient/ATC class (drugs) levels. Other demographic measures include comorbidity risk scores (Charlson, DCSI (diabetes complications severity index), CHADS2 (congestive heart failure, hypertension, age, diabetes and stroke 2) and CHAD2VASc (CHADS2 plus vascular disease history)). From prior work, feature counts have ranged in the 1000s–10 000s, and these large-scale PS models have outperformed high-dimensional PS (hdPS)[103] in simulation and real-world examples.[31] Given the subcutaneous route of administration of GLP1RAs compared with other drugs administered orally, device codes that represent needles and associated health management encounters will be excluded from PS construction.

We will:

- Exclude patients who have experienced the outcome prior to their index date.
- Stratify and variable-ratio match patients by PS.
- Use Cox proportional hazards models.

to estimate HRs between alternative target and comparator treatments for the risk of each outcome in each data source. In addition, we will perform a sensitivity analysis that does not exclude individuals who previously experienced a glycaemic control outcome before the index date. The regression will condition on the PS strata/matching unit with treatment allocation as the sole explanatory variable and censor patients at the end of their time-at-risk (TAR) or data source observation period. We will prefer stratification over matching if both sufficiently balance patients (see the Sample size and study power section), as the former optimises patient inclusions and thus generalisability.

We will execute each comparison using three different TAR definitions, reflecting different and important causal contrasts:

- Intent to treat (TAR: index +1 → end of observation) captures both direct treatment effects and (long-term) behavioural/treatment changes that initial assignment triggers.[104]
- On-treatment 1 (TAR: index +1 → treatment discontinuation) is more patient centred[105] and captures direct treatment effect while allowing for escalation with additional T2DM agents.
- On-treatment 2 (TAR: index +1 → discontinuation or escalation with T2DM agents) carries the least possible confounding with other concurrent T2DM agents.

Our 'on-treatment' is often called 'per-protocol'.[106] Systematically executing with multiple causal contrasts enables us to identify potential biases that missing prescription data, treatment escalation and behavioural changes introduce, while preserving the ease of intent-to-treat interpretation and power if the data demonstrate them as unbiased. Online supplemental appendix A.5 reports the modified cohort exit rule for the on-treatment-2 TAR.

We will aggregate HR estimates across non-overlapping data sources to produce meta-analytic estimates using a random-effects meta-analysis.[107] This classic meta-analysis assumes that per-data source likelihoods are approximately normally distributed.[108] This assumption fails when outcomes are rare as we expect for some safety events. Here, our recent research shows that as the number of data sources increases, the non-normality effect increases to where coverage of 95% CIs can be as low as 5%. To counter this, we will also apply a Bayesian meta-analysis model[109 110] that neither assumes normality nor requires patient-level data sharing by building on composite likelihood methods[111] and enables us to introduce appropriate overlap weights between data sources.

Residual study bias from unmeasured and systematic sources often remains in observational studies even after controlling for measured confounding through PS adjustment.[32 33] For each comparison-outcome effect, we will conduct negative control (falsification) outcome experiments, where the null hypothesis of no effect is believed to be true, using approximately 100 controls. We identified these controls through a data-rich algorithm[112] that identifies prevalent OMOP condition concept occurrences that lack evidence of association with exposures in published literature, drug–product labelling and spontaneous reports, and were then adjudicated by clinical review. We previously validated 60 of the controls in LEGEND for Hypertension (LEGEND-HTN).[22] Online supplemental appendix C lists these negative controls and their OMOP condition concept IDs.

Using the empirical null distributions from these experiments, we will calibrate each study effect HR estimate, its 95% CI and the p value to reject the null hypothesis of no differential effect.[34] We will declare an HR as significantly different from no effect when its calibrated $p < 0.05$ without correcting for multiple testing. Finally, blinded to all trial results, study investigators will evaluate study diagnostics for all comparisons to assess if they were likely to yield unbiased estimates (see the Sample size and study power section).

Sensitivity analyses and missingness

Because of the potential confounding effect of glycaemic control at baseline between treatment choice and outcomes and to better understand the impact of limited glucose level measurements on effectiveness and safety estimation that arises in administrative claims and some EHR data, we will perform prespecified sensitivity analyses for all studies within data sources that contain reliable glucose or haemoglobin A1c measurements. Within a study, for each exposure pair, we will first rebuild PS models, where we additionally include baseline glucose or haemoglobin A1c measurements as patient characteristics, stratify or match patients under the new PS models that directly adjust for potential confounding by

glycaemic control and then estimate effectiveness and safety HRs.

A limitation of the Cox model is that no doubly robust procedure is believed to exist for estimating HRs, due to their non-collapsibility.[113] Doubly robust procedures combine baseline patient characteristic-adjusted outcome and PS models to control for confounding and, in theory, remain unbiased when either (but not necessarily both) model is correctly specified.[114] Doubly robust procedures do exist for hazard differences[115] and we will validate the appropriateness of our univariable Cox modelling by comparing estimate differences under an additive hazards model[116] with and without doubly robust adjustment.[117] In practice, however, neither the outcome nor PS model is correctly specified, leading to systematic error in the observational setting.

Missing data of potential concern are patient demographics (gender, age and race) for our inclusion criteria. We will include only individuals whose baseline eligibility can be characterised that will most notably influence race subgroup assessments in the heterogeneity study. No further missing data can arise in our large-scale PS models because all features, with the exception of demographics, simply indicate the presence or absence of health records in a given time period. Finally, we limit the impact of missing data, such as prescription information, relating to exposure TAR by entertaining multiple definitions.[29] In all reports, we will clearly tabulate numbers of missing observations and patient attrition.

SAMPLE SIZE AND STUDY POWER

Within each data source, we will execute all comparisons with ≥1000 eligible patients per arm. Blinded to effect estimates, investigators and stakeholders will evaluate extensive study diagnostics for each comparison to assess reliability and generalisability, and only report risk estimates that pass.[25 35] These diagnostics will include:

1. Minimum detectable risk ratio as a typical proxy for power.
2. Preference score distributions to evaluate empirical equipoise10 and population generalisability.
3. Extensive patient characteristics to evaluate cohort balance before and after PS adjustment.
4. Negative control calibration plots to assess residual bias.
5. Kaplan-Meier plots to examine HR proportionality assumptions.

We will define cohorts to stand in empirical equipoise if the majority of patients carry preference scores between 0.3 and 0.7 and to achieve balance if all after-adjustment characteristics return absolute standardised mean differences <0.1.[118]

STRENGTHS AND LIMITATIONS

Strengths

LEGEND-T2DM is, to the best of our knowledge, the largest and most comprehensive study to provide evidence about the comparative effectiveness and safety of second-line T2DM agents. The LEGEND-T2DM studies will encompass over 1 million patients initiating second-line T2DM agents across at least 13 databases from 5 countries and will examine all pairwise comparisons between the four second-line drug classes against a panel of to-do health outcomes. Through an international network, LEGEND-T2DM seeks to take advantage of disparate health databases drawn from different sources and across a range of countries and practice settings. These large-scale and unfiltered populations better represent real-world practice than the restricted study populations in prescribed treatment and follow-up settings from RCTs. Our use of the OMOP CDM allows extension of the LEGEND-T2DM experiment to future databases and allows replication of these results on licensable databases that were used in this experiment while still maintaining patient privacy on patient-level data.

LEGEND-T2DM further advances the statistically rigorous and empirically validated methods we have developed in OHDSI that specifically address bias inherent in observational studies and allow for reliable causal inference. Patient characteristics and their treatment choices are likely to confound comparative effectiveness and safety estimates. Our approach combines active comparator new-user designs that emulate randomised clinical trials with large-scale propensity adjustment for measured confounding, a large set of negative control outcome experiments to address unmeasured and systematic bias, and full disclosure of hypotheses tested.

Each LEGEND-T2DM aim will represent evidence synthesis from a large number of bespoke studies across multiple data sources. Addressing questions one bespoke study at a time is prone to errors arising from multiple testing, random variation in effect estimates and publication bias. LEGEND-T2DM is designed to avoid these concerns through methodologic best practices[119] with full study diagnostics and external replication.

Through open science, LEGEND-T2DM will allow any interested investigators to engage as partners in our work at many levels. We will publicly develop all protocols and analytic code. This invites additional data custodians to participate in LEGEND-T2DM and enables others to modify and reuse our approach for other investigations. We will also host real-time access to all study result artefacts for outside analysis and interpretation. Such an open science framework ensures a feed-forward effect on other scientific contributions in the community. Collectively, LEGEND-T2DM will generate patient-centred, high quality, generalisable evidence that will transform the clinical management of T2DM through our active collaboration with patients, clinicians and national medical societies. LEGEND-T2DM will spur scientific innovation

Table 4 IRB approval or waiver statement from partners

Data source	Statement
IBM MarketScan CCAE	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
IBM MarketScan Medicare Supplemental Database (MDCR)	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
IBM MarketScan Multi-State Medicaid Database (MDCD)	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
IOC	This is a retrospective database study on de-identified data and is deemed not human subject research. Approval is provided for OHDSI network studies
JMDC	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
Korea NHIS	Ajou University IRB (AJIRB-MED-EXP-17–054 for LEGEND-HTN) and approval expected shortly for LEGEND-T2DM
Optum Clinformatics Data Mart (Optum)	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
CUIMC	Use of the CUIMC data source was approved by the Columbia University Institutional Review Board as an OHDSI network study (IRB# AAAO7805)
Department of Veterans Affairs (VA)	Use of the VA-OMOP data source was reviewed by the Department of Veterans Affairs Central IRB and was determined to meet the criteria for exemption under Exemption Category 4 (3) and approved the request for Waiver of HIPAA Authorisation
Information System for Research in Primary Care (SIDIAP)	Use of the SIDIAP data source was approved by the Clinical Research Ethics Committee of IDIAPJGol (project code: 20/070-PCV)
IQVIA Disease Analyzer,Germany	This is a retrospective database study on de-identified data and is deemed not human subject research. Approval is provided for OHDSI network studies
OptumEHR	New England IRB and was determined to be exempt from broad IRB approval, as this research project did not involve human subject research
YNHHS	Use of the YNHHS EHR data source was approved by the Yale University IRB as an OHDSI network study (IRB# pending)

CCAE, Commercial Claims and Encounters; CUIMC, Columbia University Irving Medical Center; HTN, hypertension; IOC, IQVIA Open Claims; IRB, institutional review board; JMDC, Japan Medical Data Center; LEGEND-T2DM, large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus; NHIS, National Health Insurance Service; OHDSI, Observational Health Data Sciences and Informatics; OMOP, Observational Medical Outcomes Partnership; OptumEHR, Optum electronic health records; YNHHS, Yale New Haven Health System.

through the generation of open-source resources in data science).

Limitations

Even though many potential confounders will be included in these studies, there may be residual bias due to unmeasured or misspecified confounders, such as confounding by indication, differences in physician characteristics that may be associated with drug choice, concomitant use of other drugs started after the index date and informative censoring at the end of the on-treatment periods. To minimise this risk, we will use methods to detect residual bias through a large number of negative and positive controls.

Ideal negative controls carry identical confounding between exposures and the outcome of interest.[120] The true confounding structure, however, is unknowable. Instead of attempting to find the elusive perfect negative control, we will rely on a large sample of controls that represent a wide range of confounding structures. If a study comparison proves to be unbiased for all negative controls, we can feel confident that it will also be unbiased for the outcome of interest. In our previous studies,[22 25 121] using the active comparator, new-user cohort design we will employ here, we have observed minimal residual bias using negative controls. This stands in stark contrast to other designs such as the (nested) case–control that tends to show large residual bias because of incomparable exposure cohorts implied by the design.[122]

Observed follow-up times are limited and variable, potentially reducing power to detect differences in effectiveness and safety and, further, misclassification of study variables is unavoidable in secondary use of health data, so it is possible to misclassify treatments, covariates and outcomes. Based on our previous successful studies on antihypertensives, we do not expect differential

misclassification, and, therefore, bias will most likely be toward the null. Finally, the EHR databases may be missing care episodes for patients due to care outside the respective health systems. Such bias, however, will also most likely be towards the null.

Finally, since our studies focus on healthcare datasets, as opposed to vital statistics datasets, the cause of the death among those suffering sudden cardiac death in the outpatient setting will not be identified as such.

ETHICS AND DISSEMINATION

LEGEND-T2DM does not involve human subjects research. The project does, however, use human data collected during routine healthcare provision. Most often the data are de-identified within data source. All data partners executing the LEGEND-T2DM studies within their data sources will have received institutional review board (IRB) approval or waiver for participation in accordance with their institutional governance prior to execution (see table 4). LEGEND-T2DM executes across a federated and distributed data network, where analysis code is sent to participating data partners and only aggregate summary statistics are returned, with no sharing of patient-level data between organisations.

Management and reporting of adverse events and adverse reactions

LEGEND-T2DM uses coded data that already exist in electronic databases. In these types of databases, it is not usually possible to link (ie, identify a potential causal association between) a particular product and medical event for any specific individual. Thus, the minimum criteria for reporting an adverse event (ie, identifiable patient, identifiable reporter, a suspect product and event) are not available and adverse events are not reportable as individual adverse event reports. The study results will be assessed for medically important findings.

Plans for disseminating and communicating study results

Open science aims to make scientific research, including its data process and software, and its dissemination, through publication and presentation, accessible to all levels of an inquiring society, amateur or professional[123] and is a governing principle of LEGEND-T2DM. Open science delivers reproducible, transparent and reliable evidence. All aspects of LEGEND-T2DM (except private patient data) will be open and we will actively encourage other interested researchers, clinicians and patients to participate. This differs fundamentally from traditional studies that rarely open their analytic tools or share all result artefacts, and inform the community about hard-to-verify conclusions at completion.

Transparent and re-usable research tools

We will publicly register this protocol and announce its availability for feedback from stakeholders, the OHDSI community and within clinical professional societies. This protocol will link to open-source code for all steps to generating diagnostics, effect estimates, figures and tables. Such transparency is possible because we will construct our studies on top of the OHDSI tool stack of open-source software tools that are community developed and rigorously tested.[25] We will publicly host LEGEND-T2DM source code at https://github.com/ohdsi-studies/LegendT2dm, allowing public contribution and review, and free re-use for anyone's future research.

Continuous sharing of results

LEGEND-T2DM embodies a new approach to generating evidence from healthcare data that overcome weaknesses in the current process of answering and publishing (or not) one question at a time. Generating evidence for thousands of research and control questions using a systematic process enables us to not only evaluate that process and the coherence and consistency of the evidence but also to avoid p-hacking and publication bias.[35] We will store and openly communicate all these results as they become available using a user-friendly web-based app that serves up all descriptive statistics, study diagnostics and effect estimates for each cohort comparison and outcome. Open access to this app will be through a public facing LEGEND-T2DM webpage.

Dissemination through scientific meetings and publications

We will deliver multiple presentations annually at scientific venues including the annual meetings of the American Diabetes Association, American College of Cardiology, American Heart Association and American Medical Informatics Association. We will also prepare multiple scientific publications for clinical, informatics and statistical journals.

Dissemination to general public

We believe in sharing our findings that will guide clinical care with the public. LEGEND-T2DM will use social media (Twitter) to facilitate this. With dedicated support from the OHDSI communications specialist, we will deliver regular press releases at key project stages, distributed via the extensive media networks of UCLA, Columbia and Yale.

PATIENT AND PUBLIC INVOLVEMENT

No patients were involved in the design of our studies.

Author affiliations
[1]Section of Cardiovascular Medine, Yale School of Medicine, New Haven, Connecticut, USA
[2]Center for Outcomes Research and Evaluation, Yale University School of Medicine, New Haven, Connecticut, USA
[3]Department of Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA
[4]Department of Biostatistics, University of California, Los Angeles, Los Angeles, California, USA
[5]Department of Biomedical Informatics, Columbia University Medical Center, New York, New York, USA
[6]Department of Translational Data Science and Informatics, Geisinger, Danville, Pennsylvania, USA

[7]New York-Presbyterian Hospital, New York, New York, USA
[8]Department of Biomathematics, University of California, Los Angeles, Los Angeles, California, USA
[9]Department of Human Genetics, University of California, Los Angeles, Los Angeles, California, USA
[10]VA Informatics and Computing Infrastructure, US Department of Veterans Affairs, Salt Lake City, Utan, USA

Twitter Harlan M Krumholz @hmkyale and Marc A Suchard @suchard_group

Contributors RK and MAS conceived the research and drafted the proposal in consultation with MS, YL, AO, RC, GH, PR and HMK, who provided critical feedback on the research proposal.

Funding This protocol was partially funded through the National Institutes of Health grants K23 HL153775, R01 LM006910 and R01 HG006139 and an Intergovernmental Personnel Act agreement with the US Department of Veterans Affairs. The funders had no role in the design and conduct of the protocol; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests This protocol is undertaken within Observational Health Data Sciences and Informatics (OHDSI), an open collaboration. RK is a founder of Evidence2Health, and receives grant funding from the US National Institutes of Health. MJS and PBR are employees of Janssen Research and Development and shareholders in John & Johnson. GH receives grant funding from the US National Institutes of Health and the US Food & Drug Administration and contracts from Janssen Research and Development. HMK receives grants from the US Food & Drug Administration, Medtronics and Janssen Research and Development, is co-founder of HugoHealth and chairs the Cardiac Scientific Advisory Board for UnitedHealth. MAS receives grant funding from the US National Institutes of Health, the US Department of Veterans Affairs and the US Food & Drug Administration and contracts from Janssen Research and Development and IQVIA.

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.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Rohan Khera http://orcid.org/0000-0001-9467-6199
Martijn J Schuemie http://orcid.org/0000-0002-0817-5361
Yuan Lu http://orcid.org/0000-0001-5264-2169
Anna Ostropolets http://orcid.org/0000-0002-0847-6682
Harlan M Krumholz http://orcid.org/0000-0003-2046-127X
Marc A Suchard http://orcid.org/0000-0001-9818-479X

REFERENCES

1 Lo C, Toyama T, Wang Y. Insulin and glucose-lowering agents for treating people with diabetes and chronic kidney disease. *Cochrane database of systematic reviews* 2018.
2 North EJ, Newman JD. Review of cardiovascular outcomes trials of sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists. *Curr Opin Cardiol* 2019;34:687–92.
3 Zinman B, Wanner C, Lachin JM, *et al*. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2015;373:2117–28.
4 Neal B, Perkovic V, Mahaffey KW, *et al*. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med* 2017;377:644–57.
5 Marso SP, Daniels GH, Brown-Frandsen K, *et al*. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2016;375:311–22.
6 Marso SP, Bain SC, Consoli A, *et al*. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2016;375:1834–44.
7 Scirica BM, Bhatt DL, Braunwald E, *et al*. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med* 2013;369:1317–26.
8 White WB, Cannon CP, Heller SR, *et al*. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. *N Engl J Med* 2013;369:1327–35.
9 Green JB, Bethel MA, Armstrong PW, *et al*. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2015;373:232–42.
10 Rosenstock J, Kahn SE, Johansen OE, *et al*. Effect of linagliptin vs glimepiride on major adverse cardiovascular outcomes in patients with type 2 diabetes: the Carolina randomized clinical trial. *JAMA* 2019;322:1155–66.
11 Cefalu WT, Kaul S, Gerstein HC, *et al*. Cardiovascular Outcomes Trials in Type 2 Diabetes: Where Do We Go From Here? Reflections From a *Diabetes Care* Editors' Expert Forum. *Diabetes Care* 2018;41:14–31.
12 Palmer SC, Tendal B, Mustafa RA, *et al*. Sodium-Glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2021;372:m4573.
13 Qiu M, Ding L-L, Wei X-B, *et al*. Comparative efficacy of glucagon-like peptide 1 receptor agonists and sodium glucose cotransporter 2 inhibitors for prevention of major adverse cardiovascular events in type 2 diabetes: a network meta-analysis. *J Cardiovasc Pharmacol* 2021;77:34–7.
14 Yamada T, Wakabayashi M, Bhalla A, *et al*. Cardiovascular and renal outcomes with SGLT-2 inhibitors versus GLP-1 receptor agonists in patients with type 2 diabetes mellitus and chronic kidney disease: a systematic review and network meta-analysis. *Cardiovasc Diabetol* 2021;20:14.
15 Puhan MA, Schünemann HJ, Murad MH, *et al*. A grade Working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630.
16 Brignardello-Petersen R, Izcovich A, Rochwerg B, *et al*. Grade approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ* 2020;371:m3907.
17 McCoy RG, Dykhoff HJ, Sangaralingham L, *et al*. Adoption of new glucose-lowering medications in the U.S.-The case of SGLT2 inhibitors: nationwide cohort study. *Diabetes Technol Ther* 2019;21:702–12.
18 Curtis HJ, Dennis JM, Shields BM, *et al*. Time trends and geographical variation in prescribing of drugs for diabetes in England from 1998 to 2017. *Diabetes Obes Metab* 2018;20:2159–68.
19 Arnold SV, Inzucchi SE, Tang F, *et al*. Real-World use and modeled impact of glucose-lowering therapies evaluated in recent cardiovascular outcomes trials: an NCDR® research to practice project. *Eur J Prev Cardiol* 2017;24:1637–45.
20 Dave CV, Schneeweiss S, Wexler DJ, *et al*. Trends in clinical characteristics and prescribing preferences for SGLT2 inhibitors and GLP-1 receptor agonists, 2013-2018. *Diabetes Care* 2020;43:921–4.
21 Le P, Chaitoff A, Misra-Hebert AD. Use of antihyperglycemic medications in US. Adults: An analysis of the national health and nutrition examination survey.. *Diabetes care* 2020;43:1227–33.
22 Suchard MA, Schuemie MJ, Krumholz HM, *et al*. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019;394:1816–26.
23 Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol* 2015;11:437–41.
24 Ryan PB, Schuemie MJ, Gruber S, *et al*. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf* 2013;36 Suppl 1:S59–72.
25 Schuemie MJ, Cepeda MS, Suchard MA, *et al*. How Confident are we about observational findings in healthcare: a benchmark study.

Harv Data Sci Rev 2020;2. doi:10.1162/99608f92.147cc28e. [Epub ahead of print: 31 01 2020].
26 Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858–68.
27 Gagne JJ, Fireman B, Ryan PB, *et al*. Design considerations in an active medical product safety monitoring system. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:32–40.
28 Johnson ES, Bartman BA, Briesacher BA, *et al*. The incident user design in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2013;22:1–6.
29 Schneeweiss S, Patrick AR, Stürmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007;45:S131–42.
30 Suissa S, Moodie EEM, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017;26:459–68.
31 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018;47:2005–14.
32 Schuemie MJ, Ryan PB, DuMouchel W, *et al*. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014;33:209–18.
33 Schuemie MJ, Hripcsak G, Ryan PB, *et al*. Robust empirical calibration of p-values using observational data. *Stat Med* 2016;35:3883–8.
34 Schuemie MJ, Hripcsak G, Ryan PB, *et al*. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A* 2018;115:2571–7.
35 Schuemie MJ, Ryan PB, Hripcsak G, *et al*. Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci* 2018;376. doi:10.1098/rsta.2017.0356. [Epub ahead of print: 13 Sep 2018].
36 Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, *et al*. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. *JAMA* 2010;304:411–8.
37 Turner RM, Kwok CS, Chen-Turner C, *et al*. Thiazolidinediones and associated risk of bladder cancer: a systematic review and meta-analysis. *Br J Clin Pharmacol* 2014;78:258–73.
38 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
39 Hernán M. *Antihyperglycemic therapy and cardiovascular risk: design and emulation of a target trial using healthcare databases*. Patient-Centered Outcomes Research Institute, 2019.
40 American Diabetes Association. 8. Pharmacologic Approaches to Glycemic Treatment: *Standards of Medical Care in Diabetes-2018*. *Diabetes Care* 2018;41:S73–85.
41 Nathan DM, Buse JB, Kahn SE, *et al*. Rationale and design of the glycemia reduction approaches in diabetes: a comparative effectiveness study (grade). *Diabetes Care* 2013;36:2254–61.
42 . 9. Pharmacologic Approaches to Glycemic Treatment: *Standards of Medical Care in Diabetes-2020*. *Diabetes Care* 2020;43:S98–110.
43 Schuemie MJ, Ryan PB, Pratt N. Large-Scale evidence generation and evaluation across a network of databases (legend): assessing validity using hypertension as a case study. *J Am Med Inform Assoc : JAMIA*:ocaa124.
44 Ryan PB, Buse JB, Schuemie MJ, *et al*. Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: a real-world meta-analysis of 4 observational databases (OBSERVE-4D). *Diabetes Obes Metab* 2018;20:2585–97.
45 You SC, Rho Y, Bikdeli B. Association of ticagrelor versus clopidogrel with net adverse clinical events in patients with acute coronary syndrome undergoing percutaneous coronary intervention in clinical practice. *J Am Med Assoc*. In Press.
46 Wang Y, Desai M, Ryan PB. Incidence of diabetic ketoacidosis among patients with type 2 diabetes mellitus treated with SGLT2 inhibitors and other antihyperglycemic agents. *Diabetes Res Clin Pract* 2017;128:83–90.
47 Weinstein RB, Ryan PB, Berlin JA, *et al*. Channeling bias in the analysis of risk of myocardial infarction, stroke, gastrointestinal bleeding, and acute renal failure with the use of paracetamol compared with ibuprofen. *Drug Saf* 2020;43:927–42.
48 Yuan Z, DeFalco FJ, Ryan PB, *et al*. Risk of lower extremity amputations in people with type 2 diabetes mellitus treated with sodium-glucose co-transporter-2 inhibitors in the USA: a retrospective cohort study. *Diabetes Obes Metab* 2018;20:582–9.
49 Ammann EM, Schweizer ML, Robinson JG. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for acute myocardial infarction (AMI) among intravenous immune globulin (IGIV) users in the sentinel distributed database. *Pharmacoepidemiol Drug Saf* 2018;27:398–404.
50 Floyd JS, Blondon M, Moore KP. Validation of methods for assessing cardiovascular disease using electronic health data in a cohort of veterans with diabetes. *Pharmacoepidemiol Drug Saf* 2016;25:467–71.
51 Rubbo B, Fitzpatrick NK, Denaxas S. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: a systematic review and recommendations. *Int J Cardiol* 2015;187:705–11.
52 Singh S, Fouayzi H, Anzuoni K. Diagnostic algorithms for cardiovascular death in administrative claims databases: a systematic review. *Drug Saf* 2018;23:018–754.
53 Wahl PM, Rodgers K, Schneeweiss S, *et al*. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf* 2010;19:596–603.
54 Normand SL, Morris CN, Fung KS, *et al*. Development and validation of a claims based index for adjusting for risk of mortality: the case of acute myocardial infarction. *J Clin Epidemiol* 1995;48:229–43.
55 Andrade SE, Harrold LR, Tjia J, *et al*. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:100–28.
56 Park TH, Choi JC. Validation of stroke and thrombolytic therapy in Korean National health insurance claim data. *J Clin Neurol* 2016;12:42–8.
57 Gon Y, Kabata D, Yamamoto K, *et al*. Validation of an algorithm that determines stroke diagnostic code accuracy in a Japanese hospital-based cancer registry using electronic medical records. *BMC Med Inform Decis Mak* 2017;17:157.
58 Sung S-F, Hsieh C-Y, Lin H-J, *et al*. Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. *Int J Cardiol* 2016;215:277–82.
59 Tu K, Wang M, Young J, *et al*. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMRALD as a reference standard. *Can J Cardiol* 2013;29:1388–94.
60 Yuan Z, Voss EA, DeFalco FJ, *et al*. Risk prediction for ischemic stroke and transient ischemic attack in patients without atrial fibrillation: a retrospective cohort study. *J Stroke Cerebrovasc Dis* 2017;26:1721–31.
61 Hennessy S, Leonard CE, Freeman CP, *et al*. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. *Pharmacoepidemiol Drug Saf* 2010;19:555–62.
62 Kaspar M, Fette G, Güder G, *et al*. Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information. *Clin Res Cardiol* 2018;107:778–87. Epub 2018 Apr 17.
63 Feder SL, Redeker NS, Jeon S, *et al*. Validation of the ICD-9 diagnostic code for palliative care in patients hospitalized with heart failure within the Veterans health administration. *Am J Hosp Palliat Care* 2018;35:959–65.
64 Rosenman M, He J, Martin J, *et al*. Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory. *J Am Med Inform Assoc* 2014;21:345–52. Epub 2013 Oct 10.
65 Voors AA, Ouwerkerk W, Zannad F, *et al*. Development and validation of multivariable models to predict mortality and hospitalization in patients with heart failure. *Eur J Heart Fail* 2017;19:627–34.
66 Floyd JS, Wellman R, Fuller S, *et al*. Use of electronic health data to estimate heart failure events in a population-based cohort with CKD. *Clin J Am Soc Nephrol* 2016;11:1954–61.
67 Gini R, Schuemie MJ, Mazzaglia G. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from italian general practitioners' electronic medical records: A validation study. *BMJ Open* 2016;6:e012413.
68 Afzal Z, Schuemie MJ, van Blijderveen JC, van BJC, *et al*. Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. *BMC Med Inform Decis Mak* 2013;13:30.

69 Lenihan CR, Montez-Rath ME, Mora Mangano CT, *et al*. Trends in acute kidney injury, associated use of dialysis, and mortality after cardiac surgery, 1999 to 2008. *Ann Thorac Surg* 2013;95:20–8.
70 Winkelmayer WC, Schneeweiss S, Mogun H, *et al*. Identification of individuals with CKD from Medicare claims data: a validation study. *Am J Kidney Dis* 2005;46:225–32.
71 Grams ME, Waikar SS, MacMahon B, *et al*. Performance and limitations of administrative data in the identification of AKI. *Clin J Am Soc Nephrol* 2014;9:682–9.
72 Arnold J, Ng KP, Sims D, *et al*. Incidence and impact on outcomes of acute kidney injury after a stroke: a systematic review and meta-analysis. *BMC Nephrol* 2018;19:283. 10.1186/s12882–018–1085–0.
73 Sutherland SM, Byrnes JJ, Kothari M, *et al*. Aki in hospitalized children: comparing the pRIFLE, akin, and KDIGO definitions. *Clin J Am Soc Nephrol* 2015;10:554–61.
74 Waikar SS, Wald R, Chertow GM, *et al*. Validity of international classification of diseases, ninth revision, clinical modification codes for acute renal failure. *J Am Soc Nephrol* 2006;17:1688–94.
75 Rhee C, Murphy MV, Li L, *et al*. Improving documentation and coding for acute organ dysfunction biases estimates of changing sepsis severity and burden: a retrospective study. *Crit Care* 2015;19:338.
76 Vashisht R, Jung K, Schuler A, *et al*. Association of hemoglobin A1c levels with use of sulfonylureas, dipeptidyl peptidase 4 inhibitors, and thiazolidinediones in patients with type 2 diabetes treated with metformin: analysis from the observational health data sciences and informatics initiative. *JAMA Netw Open* 2018;1:e181755–5.
77 Broder MS, Chang E, Cherepanov D, *et al*. Identification of potential markers for Cushing disease. *Endocr Pract* 2016;22:567–74.
78 Williams BA. The clinical epidemiology of fatigue in newly diagnosed heart failure*BMC Cardiovasc Disord* 2017;17
79 Yabe D, Kuwata H, Kaneko M, *et al*. Use of the Japanese health insurance claims database to assess the risk of acute pancreatitis in patients with diabetes: comparison of DPP-4 inhibitors with other oral antidiabetic drugs. *Diabetes Obes Metab* 2015;17:430–4.
80 Dore DD, Hussein M, Hoffman C, *et al*. A pooled analysis of exenatide use and risk of acute pancreatitis. *Curr Med Res Opin* 2013;29:1577–86.
81 Dore DD, Chaudhry S, Hoffman C, *et al*. Stratum-specific positive predictive values of claims for acute pancreatitis among commercial health insurance plan enrollees with diabetes mellitus. *Pharmacoepidemiol Drug Saf* 2011;20:209–13.
82 Chen H-J, Wang J-J, Tsay W-I, *et al*. Epidemiology and outcome of acute pancreatitis in end-stage renal disease dialysis patients: a 10-year national cohort study. *Nephrol Dial Transplant* 2017;32:1731–6.
83 Ooba N, Setoguchi S, Ando T. Claims-based definition of death in Japanese claims database: validity and implications. *PLoS One* 2013;8:e66116.
84 Robinson TE, Elley CR, Kenealy T, *et al*. Development and validation of a predictive risk model for all-cause mortality in type 2 diabetes. *Diabetes Res Clin Pract* 2015;108:482–8.
85 Wang L, Voss EA, Weaver J, *et al*. Diabetic ketoacidosis in patients with type 2 diabetes treated with sodium glucose co-transporter 2 inhibitors versus other antihyperglycemic agents: an observational study of four us administrative claims databases. *Pharmacoepidemiol Drug Saf* 2019;28:1620–8.
86 Buono JL, Mathur K, Averitt AJ, *et al*. Economic burden of irritable bowel syndrome with diarrhea: retrospective analysis of a U.S. commercially insured population. *J Manag Care Spec Pharm* 2017;23:453–60.
87 Krishnarajah G, Duh MS, Korves C, *et al*. Public health impact of complete and incomplete rotavirus vaccination among commercially and Medicaid insured children in the United States. *PLoS One* 2016;11:e0145977. eCollection 2016.
88 Panozzo CA, Becker-Dreps S, Pate V, *et al*. Direct, indirect, total, and overall effectiveness of the rotavirus vaccines for the prevention of gastroenteritis hospitalizations in privately insured US children, 2007-2010. *Am J Epidemiol* 2014;179:895–909.
89 Nichols GA, Brodovicz KG, Kimes TM, *et al*. Prevalence and incidence of urinary tract and genital infections among patients with and without type 2 diabetes. *J Diabetes Complications* 2017;31:1587–91.
90 Abbas S, Ihle P, Harder S, *et al*. Risk of hyperkalemia and combined use of spironolactone and long-term ACE inhibitor/angiotensin receptor blocker therapy in heart failure using real-life data: a population- and insurance-based cohort. *Pharmacoepidemiol Drug Saf* 2015;24:406–13.
91 Betts KA, Woolley JM, Mu F, *et al*. The prevalence of hyperkalemia in the United States. *Curr Med Res Opin* 2018;34:971–8.
92 Fitch K, Woolley JM, Engel T, *et al*. The clinical and economic burden of hyperkalemia on Medicare and commercial payers. *Am Health Drug Benefits* 2017;10:202–10.
93 Leonard CE, Han X, Brensinger CM, *et al*. Comparative risk of serious hypoglycemia with oral antidiabetic monotherapy: a retrospective cohort study. *Pharmacoepidemiol Drug Saf* 2018;27:9–18.
94 Chrischilles E, Rubenstein L, Chao J, *et al*. Initiation of nonselective alpha1-antagonist therapy and occurrence of hypotension-related adverse events among men with benign prostatic hyperplasia: a retrospective cohort study. *Clin Ther* 2001;23:727–43.
95 Goldstein JL, Zhao SZ, Burke TA. Incidence of outpatient physician claims for upper gastrointestinal symptoms among new users of celecoxib, ibuprofen, and naproxen in an insured population in the United States. *Am J Gastroenterol*;2003:2627–34.
96 Donga PZ, Bilir SP, Little G, *et al*. Comparative treatment-related adverse event cost burden in immune thrombocytopenic purpura. *J Med Econ* 2017;20:1200–6.
97 Marrett E, Kwong WJ, Frech F, *et al*. Health care utilization and costs associated with nausea and vomiting in patients receiving oral immediate-release opioids for outpatient acute pain management. *Pain Ther* 2016;5:215–26. Epub 2016 Oct 4.
98 Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:154–62.
99 Burwen DR, Wu C, Cirillo D, *et al*. Venous thromboembolism incidence, recurrence, and mortality based on women's health Initiative data and Medicare claims. *Thromb Res* 2017;150:78–85.
100 Coleman CI, Peacock WF, Fermann GJ, *et al*. External validation of a multivariable claims-based rule for predicting in-hospital mortality and 30-day post-pulmonary embolism complications. *BMC Health Serv Res* 2016;16:610.
101 Ammann EM, Cuker A, Carnahan RM, *et al*. Chart validation of inpatient International classification of diseases, ninth revision, clinical modification (ICD-9-CM) administrative diagnosis codes for venous thromboembolism (VTe) among intravenous immune globulin (IGIV) users in the sentinel distributed database. *Medicine* 2018;97:e9960.
102 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
103 Schneeweiss S, Rassen JA, Glynn RJ, *et al*. High-Dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009;20:512–22.
104 Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9:48–55.
105 Murray EJ, Caniglia EC, Swanson SA, *et al*. Patients and Investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. *J Clin Epidemiol* 2018;103:10–21.
106 Hernán MA, Robins JM. Per-Protocol analyses of pragmatic trials. *N Engl J Med* 2017;377:1391–8.
107 DerSimonian R, Laird N. Meta-Analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
108 Gronsbell J, Hong C, Nie L, *et al*. Exact inference for the random-effect model for meta-analyses with rare events. *Stat Med* 2020;39:252–64.
109 Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137–59.
110 et alSchuemie MJ, Chen Y, Madigan D. Combining COX regressions across a heterogeneous distributed research network facing small and zero counts, 2021. Available: http://arxiv.org/abs/2101.01551
111 Varin C, Reid N, Firth D. An overview of composite likelihood methods. *Stat Sin* 2011.
112 Voss EA, Boyce RD, Ryan PB, *et al*. Accuracy of an automated knowledge base for identifying drug adverse reactions. *J Biomed Inform* 2017;66:72–81.
113 Dukes O, Martinussen T, Tchetgen Tchetgen EJ, *et al*. On doubly robust estimation of the hazard difference. *Biometrics* 2019;75:100–9.
114 Funk MJ, Westreich D, Wiesen C, *et al*. Doubly robust estimation of causal effects. *Am J Epidemiol* 2011;173:761–7.
115 Martinussen T, Vansteelandt S, Gerster M, *et al*. Estimation of direct effects for survival data by using the Aalen additive hazards model. *J R Stat Soc Series B Stat Methodol* 2011;73:773–88.
116 Aalen OO. A linear regression model for the analysis of life times. *Stat Med* 1989;8:907–25.
117 Wang Y, Lee M, Liu P, *et al*. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. *Epidemiology* 2017;28:771–9.

118 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009;28:3083–107.
119 Schuemie MJ, Ryan PB, Pratt N. Large-Scale evidence generation and evaluation across a network of databases (legend): principles and methods. *J Am Med Inform Assoc*:ocaa103.
120 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383–8.
121 Hripcsak G, Suchard MA, Shea S, *et al*. Comparison of cardiovascular and safety outcomes of chlorthalidone vs hydrochlorothiazide to treat hypertension. *JAMA Intern Med* 2020;180:542.
122 Schuemie MJ, Ryan PB, Man KKC, *et al*. A plea to stop using the case-control design in retrospective database studies. *Stat Med* 2019;38:4199–208.
123 Woelfle M, Olliaro P, Todd MH. Open science is a research accelerator. *Nat Chem* 2011;3:745–8.