Comparative Effectiveness of Second-line Antihyperglycemic Agents for Cardiovascular Outcomes: A Large-scale, Multinational, Federated Analysis of the LEGEND-T2DM Study

Rohan Khera MD,1,2,3 Arya Aminorroaya MD,1, Lovedeep Singh Dhingra MBBS,1 Phyllis M Thangaraj MD,1 Aline Pedroso Camargos MS,1 Fan Bu PhD,4 Xiyu Ding MS,5 Akihiko Nishimura PhD,5 Tara V Anand BS,6 Faaizah Arshad BS,7 Clair Blacketer MPH,8 Yi Chai PhD,9 Shounak Chattopadhyay PhD,7 Michael Cook BSc,5 David A Dorr MD,10 Talita Duarte-Salles PhD,11,12 Scott L DuVall PhD,13,14 Thomas Falconer MS,6 Tina E French RN,15,16 Elizabeth H Hanchrow MSN,15,16 Guneet Kaur,17 Wallis CY Lau PhD,18,19,20,21 Jia Li MS,22 Kelly Li BS,7 Yuntian Liu MPH,1,2, Yuan Lu ScD,1, Kenneth KC Man PhD,18,19,20,21 Michael E Matheny MD,15,16 Nestoras Mathioudakis MD,23 Jody-Ann McLeggon MPH,6 Michael F McLemore RN,15,16, Evan Minty MD,24 Daniel R Morales MD,17 Paul Nagy PhD,23 Anna Ostropolets PhD,8 Andrea Pistillo MSc,11 Thanh-Phuc Phan MBA,25, Nicole Pratt PhD,26, Carlen Reyes MD,11, Lauren Richter MD,6, Joseph Ross MD,27,2 Benjamin Ruan MD,6 Sarah L Seager BS,28, Katherine R Simon AA,15,16, Benjamin Viernes PhD,15,14, Jianxiao Yang MS,29, Can Yin MS,30, Seng Chan You MD,31,32, Jin J Zhou PhD,7,33, Patrick B Ryan PhD,6, Martijn J Schuemie PhD,34, Harlan M Krumholz MD,2,35, George Hripcsak MD,6, Marc A Suchard PhD,7,36,37,13

1Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University, New Haven, CT, 06510, USA
2Center for Outcomes Research and Evaluation (CORE), Yale New Haven Hospital, New Haven, CT, 06510, USA
3Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, 06510, USA
4Department of Biostatistics, University of Michigan - Ann Arbor, Ann Arbor, MI, 48105, USA
5Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, 21205, USA
6Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, 10027, USA
7Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, 90095, USA
8Observational Health Data Analytics, Janssen Research and Development, LLC, Titusville, NJ, 8560, USA
9Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, The University of Hong Kong
10Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA
11Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDiAPJGol), Barcelona, 8007, Spain
12Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
13Veterans Affairs Informatics and Computing Infrastructure, United States Department of Veterans Affairs, Salt Lake City, UT, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
14Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA
15Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN, USA
16Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
17Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, DD1 9SY, United Kingdom
18Research Department of Practice and Policy, School of Pharmacy, University College London, London, WC1H 9JP, United Kingdom
19Centre for Medicines Optimisation Research and Education, University College London Hospitals NHS Foundation Trust, London, United Kingdom
20Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
21Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, Hong Kong, Hong Kong
22Data Transformation, Analytics, and Artificial Intelligence, Real World Solutions, IQVIA, Durham, NC, USA
23Division of Endocrinology, Diabetes, and Metabolism, Johns Hopkins University School of Medicine, Baltimore, MD, USA
24Faculty of Medicine, O'Brien Institute for Public Health, University of Calgary, Calgary, AB, T2N4N1, Canada
25School of Pharmacy, Taipei Medical University
26Quality Use of Medicines and Pharmacy Research Centre, UniSA Clinical and Health Sciences, University of South Australia, Adelaide, Australia
27Section of General Medicine and National Clinician Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
28Data Transformation, Analytics, and Artificial Intelligence, Real World Solutions, IQVIA, London, UK
29Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90095, USA
30Data Transformation, Analytics, and Artificial Intelligence, Real World Solutions, IQVIA, Shanghai, China
31Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea
32Institute for Innovation in Digital Healthcare, Yonsei University College of Medicine, Seoul, South Korea
33Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90024, USA
34Epidemiology, Office of the Chief Medical Officer, Johnson & Johnson, Titusville, NJ, 8560, USA
35Section of Cardiovascular Medicine, Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, 06510, USA
36 Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
37 Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

Correspondence to:
Marc A Suchard, MD, PhD
Departments of Biomathematics, Biostatistics and Human Genetics
6558 Gonda Building, 695 Charles E. Young Drive, South Los Angeles, CA 90095-1766
(310) 825-7442; msuchard@ucla.edu; @suchard_group

Word count: 3496
Number of references: 69
Number of figures/tables: 5
ABSTRACT

Background: SGLT2 inhibitors (SGLT2is) and GLP-1 receptor agonists (GLP1-RAs) reduce major adverse cardiovascular events (MACE) in patients with type 2 diabetes mellitus (T2DM). However, their effectiveness relative to each other and other second-line antihyperglycemic agents is unknown, without any major ongoing head-to-head trials.

Methods: Across the LEGEND-T2DM network, we included ten federated international data sources, spanning 1992-2021. We identified 1,492,855 patients with T2DM and established cardiovascular disease (CVD) on metformin monotherapy who initiated one of four second-line agents (SGLT2is, GLP1-RAs, dipeptidyl peptidase 4 inhibitor [DPP4is], sulfonylureas [SUs]). We used large-scale propensity score models to conduct an active comparator, target trial emulation for pairwise comparisons. After evaluating empirical equipoise and population generalizability, we fit on-treatment Cox proportional hazard models for 3-point MACE (myocardial infarction, stroke, death) and 4-point MACE (3-point MACE + heart failure hospitalization) risk, and combined hazard ratio (HR) estimates in a random-effects meta-analysis.

Findings: Across cohorts, 16·4%, 8·3%, 27·7%, and 47·6% of individuals with T2DM initiated SGLT2is, GLP1-RAs, DPP4is, and SUs, respectively. Over 5·2 million patient-years of follow-up and 489 million patient-days of time at-risk, there were 25,982 3-point MACE and 41,447 4-point MACE events. SGLT2is and GLP1-RAs were associated with a lower risk for 3-point MACE compared with DPP4is (HR 0·89 [95% CI, 0·79-1·00] and 0·83 [0·70-0·98]), and SUs (HR 0·76 [0·65-0·89] and 0·71 [0·59-0·86]). DPP4is were associated with a lower 3-point MACE risk versus SUs (HR 0·87 [0·79-0·95]). The pattern was consistent for 4-point MACE for the comparisons above. There were no significant differences between SGLT2is and GLP1-RAs for 3-point or 4-point MACE (HR 1·06 [0·96-1·17] and 1·05 [0·97-1·13]).

Interpretation: In patients with T2DM and established CVD, we found comparable cardiovascular risk reduction with SGLT2is and GLP1-RAs, with both agents more effective than
DPP4is, which in turn were more effective than SUIs. These findings suggest that the use of GLP1-RAs and SGLT2is should be prioritized as second-line agents in those with established CVD.

**Funding:** National Institutes of Health, United States Department of Veterans Affairs

**Keywords:** Diabetes Mellitus, Type 2; Hypoglycemic Agents; Glucagon-Like Peptide-1 Receptor Agonists; Sodium-Glucose Transporter 2 Inhibitors; Comparative Effectiveness Research; Cardiovascular Diseases
RESEARCH IN CONTEXT

Evidence before this study
Sodium-glucose co-transporter-2 inhibitors (SGLT2is) and glucagon-like peptide-1 receptor agonists (GLP1-RAs) exert cardiovascular benefits beyond blood glucose control, transforming the landscape of treating type 2 diabetes mellitus (T2DM) over the past decade. Large randomized clinical trials have substantiated the cardiovascular benefits of SGLT2is and GLP1-RAs in decreasing major adverse cardiovascular events. However, trials of SGLT2is and GLP1-RAs have not been designed as head-to-head comparisons with each other or older antihyperglycemic agents such as dipeptidyl peptidase-4 inhibitors (DPP4is) and sulfonylureas (SUs). Therefore, the comparative cardiovascular effectiveness of these antihyperglycemic agents is not established. Recent work also shows a substantial proportion of patients with T2DM are often initiated on DPP4is and SUs in the absence of information on their cardiovascular effectiveness. These knowledge gaps in evidence challenge the development of treatment recommendations based on cardiovascular risk profiles in T2DM.

Added value of this study
The Large-scale Evidence Generation and Evaluation across a Network of Databases for Type 2 Diabetes Mellitus (LEGEND-T2DM) initiative is a large-scale, systematic, federated evaluation of individuals with T2DM across multiple international observational data sources, where we leveraged robust state-of-the-art methodological and analytic strategies to minimize residual confounding, publication bias, and p-hacking. In this largest study of 1.5 million patients with T2DM and established cardiovascular disease (CVD) across ten cohorts from four countries initiating a second-line antihyperglycemic agent — SGLT2is, GLP1-RAs, DPP4is, or SUs — after metformin monotherapy, there was no significant difference in the cardiovascular effectiveness of SGLT2is and GLP1-RAs. These agents are associated with an 11% and 17% lower risk of cardiovascular events compared with DPP4is, respectively. All three agents were associated with lower cardiovascular events than SUs, with 24% and 28% lower risk with
SGLT2is and GLP1-RAs, and 12% lower risk with DPP4is. These patterns were consistent across both ischemic events, such as MI and stroke, as well as for heart failure.

Implications of all the available evidence

Our study defines strong empirical evidence in support of contemporary clinical practice guidelines that suggest SGLT2is and GLP1-RAs are used for those with T2DM and CVD. Our observations also provide support for exclusive second-line use of SGLT2is and GLP1-RAs among those with CVD.
INTRODUCTION

Over the past decade, the therapeutic options for type 2 diabetes mellitus (T2DM) have undergone a significant transformation. Sodium-glucose co-transporter-2 inhibitors (SGLT2is) and glucagon-like peptide-1 receptor agonists (GLP1-RAs) have expanded the role of antihyperglycemic agents from managing high blood glucose to addressing the elevated cardiovascular risk in patients with T2DM. In several large randomized clinical trials (RCTs), SGLT2is and GLP1-RAs reduced major adverse cardiovascular events (MACE), such as myocardial infarction (MI), hospitalization for heart failure, and cardiovascular mortality. While older antihyperglycemic agents, like sulfonylureas (SUs), have not undergone similarly comprehensive trials to evaluate their cardiovascular efficacy or safety, some studies suggested their neutral effect on cardiovascular outcomes and risk of hypoglycemia. Furthermore, direct comparisons of SGLT2is and GLP1-RAs with dipeptidyl peptidase-4 inhibitors (DPP4is), which are antihyperglycemic agents with neutral effects on MACE, have not been conducted and no major trials are in progress. Nevertheless, DPP4is and SUs continue to be used in clinical practice and are recommended as second-line T2DM agents in national clinical practice guidelines.

Despite the availability of evidence from SGLT2i and GLP1-RA trials, several gaps in evidence challenge the development of treatment recommendations in T2DM. Specifically, trials of SGLT2is and GLP1-RAs were not designed as head-to-head comparisons with older agents but rather as additive treatments on top of commonly used T2DM agents. As a result, the relative cardiovascular efficacy of newer versus older agents is not known. Moreover, the comparative cardiovascular efficacy of SGLT2is and GLP1-RAs have not been evaluated. Thus far, comparative effectiveness assessments have been based on indirect estimates from clinical trials or comparative effectiveness drawn from a limited number of data sources. The findings from observational studies using single data sources are challenged by our observations that the uptake of these agents, and therefore, the selective pressures on the use
of these agents, varies considerably across sites,\textsuperscript{21} with the potential to introduce bias in the evaluation of comparative effectiveness. These evidence gaps pose a significant challenge in designing treatment algorithms that rely on the comparative effectiveness and safety of drugs.\textsuperscript{24} As a result, there is a large variation in clinical practice guidelines and clinical practice with regard to these medications, with many patients initiated on the newer therapies and many others treated with older regimens.\textsuperscript{21,29}

To address this, we conducted the Large-scale Evidence Generation and Evaluation across a Network of Databases for Type 2 Diabetes Mellitus (LEGEND-T2DM) initiative,\textsuperscript{30} a large-scale, systematic, federated evaluation of 4.7 million patients with T2DM across multiple international observational data sources, where we leveraged robust state-of-the-art methodological and analytic strategies to minimize residual confounding, publication bias, and p-hacking. Here, we compared the cardiovascular effectiveness of four second-line antihyperglycemic agents — SGLT2is, GLP1-RAs, DPP4is, or SUs — when initiated on the background of metformin therapy in T2DM.

METHODS

Data Sources

In this study, we included ten real-world data sources from the LEGEND-T2DM network, including six administrative claims and four EHR databases across four countries during 1992-2021 (figure 1). These represent data from six national-level and four health-system datasets from the US, as well as data sources from Germany, Spain, and the United Kingdom. The vast majority of patient records span from the mid-2000s to today, covering two decades of T2DM treatment as well as the introduction of many second-line antihyperglycemic agents. All LEGEND-T2DM data sources were previously standardized to the Observational Health Data Sciences and Informatics (OHDSI)’s Observational Medical Outcomes Partnership (OMOP) common data model (CDM) version 5,\textsuperscript{31} which mapped international coding systems into...
standard vocabulary concepts. The use of the OMOP CDM allows a federated analysis of data, without patient-level data sharing, using consistent cohort definitions and study design. All data partners received institutional approval or exemption for their participation. Details of data sources are presented in appendix pp 17-19.

Study Design
We identified patients with T2DM and established cardiovascular disease (CVD) on metformin monotherapy who initiated on any of the drug ingredients within one of the SGLT2i, GLP1-RA, DPP4i, or SU drug classes (appendix pp 2-10, 20). Each exposure cohort thus consists of new users of each drug class. We require patients with T2DM and CVD to have at least one year of prior observation in the database, with at least 3 months of prior metformin use before initiating a second-line agent, and no prior exposure to a comparator second-line or other antihyperglycemic agent or more than 30 days of insulin exposure (appendix pp 2-10). To ensure statistical power, we executed analyses for comparisons and data sources with at least 1,000 patients in each arm. To evaluate the comparative effectiveness of antihyperglycemic agents, we constructed target–comparator–database combinations, where we compared one antihyperglycemic agent (target) with another agent (comparator) across data sources. For this, we employed a new-user cohort design for target and comparator agents within each data source in order to emulate the hypothetical target trial. Methodological principles in the study design have been carefully constructed based on the evidence by experts and were leveraged previously to minimize bias and improve reproducibility.

Study Outcomes
Across all data sources and pairwise exposure cohorts, we assess the relative risks of two primary and four secondary cardiovascular outcomes. The two primary outcomes of interest are (1) 3-point MACE, including acute MI, stroke, and sudden cardiac death, and (2) 4-point MACE,
which additionally includes hospitalization for heart failure. Secondary outcomes of interest include the four individual MACE components. We construct outcome cohorts based on previously developed phenotypes validated and tested in prior work (appendix pp 11-14).³⁴–³⁹

For each outcome cohort, we included patients with no events prior to treatment initiation and defined continuous drug exposure as consecutive drug prescriptions with less than 30-day prescription gaps. We considered an on-treatment time-at-risk (TAR) definition that follows a patient from treatment initiation to treatment discontinuation, which captures direct treatment effects while allowing for escalation with additional T2DM agents.

**Statistical Analysis**

We employed a systematic federated analytic framework to address residual confounding, publication bias, and p-hacking.³⁰,⁴⁰ This framework uses data-driven, large-scale propensity score (PS) adjustment for measured confounding,⁴¹ a large set of negative control outcome experiments to address unmeasured confounding and systematic bias,⁴²–⁴⁴ study diagnostics to ensure validity and generalizability,⁴⁴,⁴⁵ and a principled meta-analysis approach to aggregate evidence across data sources. We used standardized vocabularies to construct consistent computable definitions of all study cohorts, covariates, and outcomes. We provide full disclosure of all hypotheses investigated and pre-specify and report all analytical procedures in the published protocol of LEGEND-T2DM.³⁰ To promote open science and avoid publication bias, we have disseminated all results in a publicly available R ShinyApp (https://data.ohdsi.org/LegendT2dmClassEvidenceExplorer/), and all analytic code is publicly available on GitHub (https://github.com/ohdsi-studies/LegendT2dm).

Within each data source, we estimated the relative risks of all six outcomes between each pair of new-user cohorts, taking one exposure cohort as the target and the other as a comparator. For each pairwise comparison and each data source, we adjusted for measured confounding and improved balance between cohorts through both matching and stratifying on
PS. We estimated the PS by a data-driven approach that adjusts for a broad range of predefined baseline patient characteristics through regularized regression. These characteristics included demographics, comorbidities, concomitant medication use, and healthcare utilization in the period before the initiation of the second-line antihyperglycemic agent. The choice of PS stratification vs. matching was based on the approach that achieves a standardized mean difference (SMD) <0·15 across all covariates. When both stratification and matching provide sufficient balance, we preferred stratification over matching and thus reported results based on stratification when available, as the former improves patient inclusion and, therefore, generalizability. We then used Cox proportional hazards models to estimate hazard ratios (HRs) of each outcome for each comparison, conditional on PS stratification or variable-ratio patient matching.

We also sought to address residual bias, which may persist in observational studies even after PS adjustment that controls for measured confounding. For this, we conducted negative control (falsification) outcome experiments for each comparison and outcome, where the null hypothesis of no differential effect (i.e., HR = 1) is believed to be true for each outcome. We selected 100 negative controls through a data-driven algorithm that identifies OMOP condition concept occurrences with similar prevalence to the outcomes of interest that lack evidence of association with exposures in published literature, drug–product labels, and spontaneous reports, which we then confirmed by expert review (appendix p 15). We list these negative controls in appendix pp 21-22. From these negative control experiments, we learn an empirical null distribution that informs residual study bias, i.e., a deviation from the empirical null across all outcomes represents a quantitative surrogate for the residual bias. We calibrate each original HR estimate to compute a calibrated HR estimate and 95% confidence interval (CI). We declare an HR as significantly different from the null if the calibrated p-value is <0·05 without considering multiple testing corrections.
We assessed, while blinded to the results, study diagnostics to ensure reliability and generalizability for all comparisons and only report estimates that pass the diagnostics. These study diagnostics included (1) minimum detectable risk ratio (MDRR) as a metric for statistical power, (2) preference score distributions between the target and comparator cohorts to evaluate empirical equipoise and population generalizability, and (3) cohort balance before and after propensity score (PS) adjustment, defined by the absolute SMDs on extensive patient characteristics. A study passed diagnostics if MDRR was less than 4, and >25% of patients had a preference score between 0.3 and 0.7 on both arms and maximum SMD <0.15 after PS adjustment. Additional diagnostics for visual examination included (4) calibration plots on negative control outcomes to examine residual bias, and (5) Kaplan-Meier plots to check proportionality assumptions for the Cox models.

We reported all HR estimates, their 95% CIs, and p-values post-calibration for studies that passed diagnostics. To aggregate evidence across non-overlapping data sources, we combined all calibrated HR estimates for each comparison using a random effects meta-analysis approach.

**RESULTS**

**Cohort Characteristics**

Across ten federated longitudinal data sources from four countries, we identified 4,730,887 patients with T2DM. This included 1,492,855 patients with T2DM and established CVD on metformin monotherapy who initiated one of the four second-line antihyperglycemic agents and had no prior use of any other antihyperglycemic agents (figure 1, appendix p 17). Among these patients, 244,694 (16.4%) initiated SGLT2is, 123,991 (8.3%) GLP1-RAs, 413,236 (27.7%) DPP4is, and 710,934 (47.6%) SUs (table 1). US Open Claims (USOC) with 1,163,413 patients with T2DM and established CVD, followed by Optum Clininformatics Extended Data Mart - Date of Death (OCEDM) (N=80,159) and Optum de-identified Electronic Health Record Dataset
(OEHR) (N=66,417) contributed the most patients to the study population. Median on-treatment time-at-risk for patients varied by drug class and database between 2.0 and 23.2 months. At least 25% of the patients were exposed to their first drug class for more than 12 months across a majority of databases (table 1).

Addressing Confounding Across Target-Comparator-Database Combinations

PS adjustment achieved pre-specified covariate balance in patient baseline characteristics for pairwise class comparisons across all databases (figure 2, appendix pp 23-108). The maximum SMD across target–comparator–database combinations consistently decreased after PS stratification (figure 2). For example, in comparison of patients initiating SGLT2is (target) with patients initiating DPP4is (comparator) in the IBM MarketScan Commercial Claims and Encounters (CCAE) database, before PS adjustment, patients initiating SGLT2is were more frequently men and had obesity and heart failure relative to patients initiating DPP4is. However, after PS adjustment, the SGLT2is or DPP4is populations were well balanced on all demographic and clinical patient characteristics (appendix pp 33-34).

Empirical Equipoise Across Target-Comparator-Database Combinations

For most data sources, all executed class comparisons were in empirical equipoise (>25% of patients had preference scores between 0.3 and 0.7 on both arms) (appendix pp 146-155). GDA, SIDIAP, and VA databases showed less equipoise for comparisons involving SGLT2is and DPP4is. However, in general, PS adjustment achieved sufficient covariate balance in terms of preference score distribution to reduce concerns that measured the estimated effects of baseline confounding biases (appendix pp 156-175). Furthermore, the study had limited residual systematic error, where calibration of effect estimates using negative control outcomes resulted in an increase in the proportion of nominal 95% CIs that included 1 for control outcomes across a majority of comparisons and databases (appendix pp 176-195). For
example, for SGLT2is vs DPP4is comparison in SIDIAP using PS matching, before calibration, the nominal CIs covered 75.0% of control estimates; after calibration, they covered 91.7% (appendix p 192).

**Comparative Effectiveness for Primary Endpoints**
During 1,337,809 patient-years of follow-up, there were 25,982 3-point MACE and 41,447 4-point MACE events. SGLT2is and GLP1-RAs were associated with a lower hazard of 3-point MACE compared with DPP4is (HR 0.89 [95% CI, 0.79-1.00] and 0.83 [0.70-0.98], respectively), and with even lower HRs of 0.76 (95% CI, 0.65-0.89) and 0.72 (95% CI, 0.58-0.88) versus SUs, respectively. DPP4is were associated with a lower risk of 3-point MACE risk versus SUs (0.87 [0.79-0.95]). A consistent pattern was observed for 4-point MACE across the above-mentioned comparisons. In direct comparisons of SGLT2is and GLP1-RAs, there were no differences in either 3-point or 4-point MACE (1.06 [0.96-1.17], and 1.05 [0.97-1.13], respectively) (appendix pp 109-121, figure 3).

**Comparative Effectiveness for Secondary Endpoints**

*Acute MI*
There were 13,536 episodes of acute MI across data sources during the follow-up. SGLT2is, GLP1-RAs, and DPP4is were associated with a lower hazard of acute MI compared with SUs (0.81 [0.69-0.95], 0.70 [0.56-0.87], and 0.86 [0.77-0.95], respectively). While GLP1-RAs were associated with a lower risk of AMI compared with DPP4is (0.85 [0.73-0.99]), the hazard of acute MI was comparable for SGLT2is vs DPP4is (0.95 [0.83-1.08]). Compared with GLP1-RAs, SGLT2is were associated with a higher hazard of acute MI (1.19 [1.05-1.35]) (appendix pp 109, 122-127, figure 3).

*Stroke*
There were 13,999 episodes of stroke events across data sources during the follow-up. SGLT2is, GLP1-RAs, and DPP4is were associated with a lower hazard of stroke compared with SUs (0.77 [0.65-0.92], 0.79 [0.63-0.98], and 0.90 [0.81-0.99], respectively). Compared with DPP4is, SGLT2is and GLP1-RAs had a comparable hazard of stroke (0.89 [0.74-1.07] and 0.89 [0.77-1.03], respectively). There was no significant difference between SGLT2is and GLP1-RAs (0.93 [0.82-1.05]) regarding the hazard of stroke (appendix pp 109, 128-133, figure 3).

**SCD**

Over follow-up, 3,789 SCD occurred across data sources. SGLT2is and GLP1-RAs were associated with a lower hazard of SCD compared with DPP4is (0.62 (0.51-0.75), and 0.65 [0.44-0.96], respectively) and SUs (0.60 [0.48-0.73], and 0.57 [0.42-0.78], respectively). DPP4is versus SUs and SGLT2is versus GLP1-RAs had comparable hazards of SCD (0.96 [0.81-1.13], and 1.11 [0.84-1.45], respectively) (appendix pp 109, 134-139, figure 3).

**HF Hospitalization**

During the follow-up, 30,743 HF hospitalizations were recorded across all databases. SGLT2is and GLP1-RAs were associated with a lower risk of HF hospitalization compared with DPP4is (0.80 [0.68-0.95] and 0.76 [0.67-0.86], respectively) and SUs (0.68 [0.58-0.80] and 0.64 [0.52-0.78], respectively). DPP4is were associated with a lower hazard of HF hospitalization compared with SUs (0.90 [0.82-0.99]). Compared with GLP1-RAs, SGLT2is had a comparable hazard for HF hospitalization (1.02 [0.93-1.12]) (appendix pp 109, 140-145, figure 3).

**Sensitivity Analyses**

Across all outcomes, 98.4% (690/701) of calibrated relative risk estimates remained within the confidence intervals of the primary meta-analysis when systematically removing each data source from the leave-one-out meta-analysis. Within the primary endpoints of 3-point MACE
and 4-point MACE and secondary endpoints of acute MI and hospitalization for heart failure, none of the leave-one-out analyses were outside of the confidence intervals of the meta-analysis (figure 4, appendix pp 196-201).

DISCUSSION

In this largest, multinational, federated study of 1.5 million patients with T2DM and established CVD across ten cohorts initiating a second-line antihyperglycemic agent after metformin monotherapy, SGLT2is and GLP1-RAs were associated with an 11% and 17% lower risk of cardiovascular events compared with DPP4is, respectively, the class of drugs approved before the US FDA-mandated cardiovascular outcomes trials for antihyperglycemic agents.50,51 All three agents were associated with lower cardiovascular events than SUs, with 24% and 28% lower risk with SGLT2is and GLP1-RAs and 12% lower risk with DPP4is. These patterns were consistent across both ischemic events, such as MI and stroke, as well as for heart failure. Between SGLT2i and GLP1-RAs, there was no difference in composite cardiovascular effectiveness outcomes – both 3-point and 4-point MACE, but there was a lower risk of AMI with GLP1-RAs compared with SGLT2i. The study accounted for multiple null control outcomes and calibrated all estimates for residual bias on null control outcomes, with multiple sensitivity analyses confirming the findings of the primary analysis.

This builds on the evidence of comparative cardiovascular effectiveness of currently used antihyperglycemic therapies, especially in the absence of head-to-head RCTs of any of these agents. There are no ongoing or future trials expected to fill this knowledge gap. A few observational studies have used single data sources to evaluate the cardiovascular outcomes associated with these drugs.52–56 Moreover, some associations observed in the present study, such as those with DPP4i, contrast with others where the smaller study population may have necessitated studying a combination of drugs directly against SU but challenges isolating an effect.57 These analyses are limited by local patterns of treatment use and challenges with
exposure and outcome ascertainment in individual sources that may confound the observed association of these agents with outcomes. Our recent work shows that the use of antihyperglycemic agents varies substantially in populations across the world.\textsuperscript{21} In the present study, we developed the target trial for head-to-head of these agents, focusing explicitly on the same disease stage, i.e. when escalation to a second-line agent is being initiated. We conducted the same study across ten different data sources, ensuring that the effects of the intervention were not limited to the selective pressures or outcome differences in a single population. We ensured the empirical equipoise by comparing SMDs before and after PS stratification. Finally, we ensured that unmeasured confounding, potentially manifesting as directional effects across several negative control outcomes, were used to calibrate the PS stratified effect estimates.

The findings have important implications for clinical practice. Current clinical practice guidelines offer the use of SGLT2is and GLP1-RAs as suggested agents to be used among those with existing cardiovascular risk.\textsuperscript{18,23,58} We find strong evidence supportive of that suggestion. However, our recent work shows a substantial proportion of patients with T2DM are often initiated on DPP4is and SUs despite their limited cardiovascular effectiveness or known superiority on other glycemic or other diabetes-related outcomes.\textsuperscript{59–61} Our observations would favor stronger support for exclusive second-line use of SGLT2is and GLP1-RAs among individuals with established CVD. There is also evidence that suggests a potentially larger reduction in risk of AMI with GLP1-RAs compared with SGLT2is, which were not replicated for stroke, questioning whether the current guideline recommendation for preferential use of GLP1-RAs in ASCVD is appropriate. Moreover, we did not observe an exclusive role of decreased HF risk with SGLT2i, as suggested in current guidelines, as there was no observed difference in HF risk among those receiving SGLT2i or GLP1-RA. The original recommendations for HF were based on inference from RCTs of individual agents and emerging evidence, such as from the STEP2-HFpEF trial that found improvements in heart failure outcomes with semaglutide support.
the potential role of GLP1-RAs on HF outcomes as well. Our study favors the continued evaluation of GLP1-RA as a therapy to reduce HF risk and improve outcomes of those with HF.

There are limitations that merit consideration. First, to enable the ability to capture information across data sources spanning administrative and EHR data sources necessitated the use of administrative definitions of both exposures and outcomes. However, to ensure the information is accurately captured, we used definitions that have been validated in prior studies. Second, we expect the identification of the exposure and outcomes to vary across data sources. However, the consistency of effect estimates across data sources further confirms the validity of the observations. Third, capturing outcomes for EHR-based datasets, such as for recurrent hospitalizations and deaths, differs from administrative sources that track events across institutions and capture out-of-hospital death events. However, we find that sources that capture cross-institution out-of-hospital information did not differ and had similar outcomes as those that focused on institutionally captured death events, further supporting the robustness of the drug association with outcomes. Fourth, we do not have information on adherence to therapies and whether differential adherence of agents would confound the observations. However, these agents are widely used without substantial adverse effects. Moreover, lack of adherence will likely bias these observations to the null, and the observed positive association across agents with known cardiovascular benefits suggests that it is not driving the observed associations. Finally, inference on all outcomes can still be subject to residual confounding. However, our use of calibrated estimates explicitly addresses the identification of residual confounding across a series of null control outcomes.

CONCLUSIONS

Our large, rigorous multinational network study of patients with T2DM and established CVD found cardiovascular risk reduction with SGLT2is and GLP1-RAs, with both agents more
effective than DPP4is, which in turn were more effective than SUs. These findings suggest that the use of GLP1-RAs and SGLT2is should be prioritized as second-line agents in those with established CVD.

**Declaration of interests**

RK is an Associate Editor at JAMA and received support from the National Heart, Lung, and Blood Institute of the National Institutes of Health (under award K23HL153775) and the Doris Duke Charitable Foundation (under award, 2022060). He also receives research support, through Yale, from Bristol-Myers Squibb, Novo Nordisk, and BridgeBio. He is a coinventor of U.S. Provisional Patent Applications 63/619,241, 63/606,203, 63/177,117, 63/346,610, 63/428,569, and 63/484,426, unrelated to current work. He is also a co-founder of Evidence2Health and Ensight-AI, both representing precision health platforms to improve evidence-based cardiovascular care. SLD reports grants from Alnylam Pharmaceuticals, Inc., AstraZeneca Pharmaceuticals LP, Biodesix, Inc, Celgene Corporation, Cerner Enviza, GSK PLC, Janssen Pharmaceuticals, Inc., Novartis International AG, Parexel International Corporation through the University of Utah or Western Institute for Veteran Research outside the submitted work. KKCM receives support from C W Maplethorpe Fellowship, National Institute of Health Research, UK, European Commission Framwork Horizon 2020, Hong Kong Research Grant Council, Innovation and Technology Commission of the Hong Kong Special Administration Region Government, outside the submitted work. DRM was supported by a Wellcome Trust Clinical Research Fellowship (214588/Z/18/Z). In the past three years, HMK received expenses and/or personal fees from UnitedHealth, Element Science, Aetna, Reality Labs, Tesseract/4Catalyst, F-Prime, the Siegfried and Jensen Law Firm, Arnold and Porter Law Firm, and Martin/Baughman Law Firm. He is a co-founder of Refactor Health and HugoHealth, and is associated with contracts, through Yale New Haven Hospital, from the Centers for Medicare & Medicaid Services and through Yale University from Johnson & Johnson. SCY
reports being a Chief Technology Officer of PHI Digital Healthcare. JRT currently receives
research support through Yale University from Johnson and Johnson to develop methods of
clinical trial data sharing, from the Food and Drug Administration for the Yale-Mayo Clinic
Center for Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938),
from the Agency for Healthcare Research and Quality (R01HS022882), and from Arnold
Ventures; formerly received research support from the Medical Device Innovation Consortium
as part of the National Evaluation System for Health Technology (NEST) and from the National
Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164,
R01HL144644); and in addition, JRT was an expert witness at the request of Relator’s
attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and
Anti-Kickback Statute against Biogen Inc. that was settled September 2022. PMT is a
coinventor of a provisional patent 63/606,203 unrelated to the current work and is funded by
5T32HL155000-03. CB, AO and PBR are employees of Johnson & Johnson and MJS is an
employee and shareholder of Johnson & Johnson. Other authors declare no conflicts of interest.

Acknowledgments

This study was partially funded through the National Institutes of Health grants K23 HL153775,
R01 LM006910, R01 HG006139 and R01 HL169954, and the US Department of Veterans
Affairs under the research priority to "Put VA Data to Work for Veterans" (VA ORD 22-D4V).
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.
REFERENCES


28 Gilbert MP, Pratley RE. GLP-1 Analogs and DPP-4 Inhibitors in Type 2 Diabetes Therapy: Review of Head-to-Head Clinical Trials. *Front Endocrinol* 2020; 11: 178.

29 Nargesi AA, Jeyashanmugaraja GP, Desai N, Lipska K, Krumholz H, Khera R. Contemporary National Patterns of Eligibility and Use of Novel Cardioprotective


### Tables and Figures

**Table 1: Population Size and Follow-up Time for Initiators of SGLT2is, GLP1-RAs, DPP4is, and SUs Across Databases.**

<table>
<thead>
<tr>
<th>Drug/Data Source</th>
<th>Patients, No.</th>
<th>On-treatment Follow-up Time (days), Median [IQR]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SGLT2i</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCAE</td>
<td>10,135</td>
<td>161 [62 - 537]</td>
</tr>
<tr>
<td>GDA</td>
<td>4,337</td>
<td>181 [97 - 537]</td>
</tr>
<tr>
<td>MDCD</td>
<td>1,314</td>
<td>114 [30 - 383]</td>
</tr>
<tr>
<td>MDCR</td>
<td>2,013</td>
<td>112 [34 - 541]</td>
</tr>
<tr>
<td>OCEDM</td>
<td>11,597</td>
<td>118 [55 - 423]</td>
</tr>
<tr>
<td>OEHR</td>
<td>8,996</td>
<td>71 [29 - 181]</td>
</tr>
<tr>
<td>SIDIAP</td>
<td>2,281</td>
<td>322 [103 - 895]</td>
</tr>
<tr>
<td>USOC</td>
<td>196,710</td>
<td>143 [61 - 525]</td>
</tr>
<tr>
<td>VA</td>
<td>7,311</td>
<td>135 [71 - 443]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>244,694</strong></td>
<td></td>
</tr>
<tr>
<td><strong>GLP1-RA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCAE</td>
<td>7,446</td>
<td>116 [45 - 441]</td>
</tr>
<tr>
<td>MDCR</td>
<td>1,120</td>
<td>94 [29 - 443]</td>
</tr>
<tr>
<td>OCEDM</td>
<td>6,546</td>
<td>98 [36 - 368]</td>
</tr>
<tr>
<td>OEHR</td>
<td>4,525</td>
<td>60 [29 - 173]</td>
</tr>
<tr>
<td>USOC</td>
<td>104,354</td>
<td>112 [52 - 432]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>123,991</strong></td>
<td></td>
</tr>
<tr>
<td><strong>DPP4i</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCAE</td>
<td>22,801</td>
<td>159 [63 - 597]</td>
</tr>
<tr>
<td>GDA</td>
<td>10,428</td>
<td>221 [97 - 741]</td>
</tr>
<tr>
<td>IMRD</td>
<td>4,024</td>
<td>354 [122 - 1,174]</td>
</tr>
<tr>
<td>MDCD</td>
<td>4,449</td>
<td>142 [56 - 555]</td>
</tr>
<tr>
<td>MDCR</td>
<td>10,541</td>
<td>187 [81 - 713]</td>
</tr>
<tr>
<td>OCEDM</td>
<td>23,918</td>
<td>150 [59 - 569]</td>
</tr>
<tr>
<td>OEHR</td>
<td>18,145</td>
<td>95 [89 - 312]</td>
</tr>
<tr>
<td>SIDIAP</td>
<td>8,126</td>
<td>695 [204 - 1,916]</td>
</tr>
<tr>
<td>USOC</td>
<td>299,835</td>
<td>157 [59 - 657]</td>
</tr>
<tr>
<td>VA</td>
<td>10,969</td>
<td>220 [90 - 707]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>413,236</strong></td>
<td></td>
</tr>
<tr>
<td><strong>SU</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCAE</td>
<td>18,093</td>
<td>148 [59 - 559]</td>
</tr>
<tr>
<td>GDA</td>
<td>1,235</td>
<td>216 [119 - 684]</td>
</tr>
<tr>
<td>IMRD</td>
<td>846</td>
<td>350 [106 - 1312]</td>
</tr>
<tr>
<td>MDCD</td>
<td>6,758</td>
<td>124 [33 - 503]</td>
</tr>
<tr>
<td>MDCR</td>
<td>8,887</td>
<td>159 [59 - 632]</td>
</tr>
<tr>
<td>OCEDM</td>
<td>38,098</td>
<td>177 [85 - 652]</td>
</tr>
<tr>
<td>OEHR</td>
<td>34,751</td>
<td>110 [89 - 386]</td>
</tr>
<tr>
<td>SIDIAP</td>
<td>619</td>
<td>600 [158 - 1,809]</td>
</tr>
<tr>
<td>USOC</td>
<td>562,514</td>
<td>183 [89 - 752]</td>
</tr>
<tr>
<td>VA</td>
<td>39,133</td>
<td>186 [90 - 689]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>710,934</strong></td>
<td></td>
</tr>
</tbody>
</table>
Abbreviations: CCAE, IBM MarketScan® Commercial Claims and Encounters Data; DPP4i, dipeptidyl peptidase 4 inhibitor; GDA, Germany Disease Analyzer; GLP1-RA, glucagon-like peptide-1 receptor agonist; IMRD, UK-IQVIA Medical Research Data; MDCD, IBM Health MarketScan® Multi-State Medicaid Database; IQR, interquartile range; MDCR, IBM Health MarketScan® Medicare Supplemental and Coordination of Benefits Database; OCEDM, Optum© Clinformatics Extended Data Mart - Date of Death; OEHR, Optum© de-identified Electronic Health Record Dataset; SGLT2i, sodium-glucose co-transporter-2 inhibitor; SIDIAP, Information System for Research in Primary Care; SU, Sulfonylurea; USOC, United States Open Claims; VA, Department of Veterans Affairs Healthcare System.
Figure 1: Study Design and Analytical Methodology to Evaluate the Comparative Effectiveness of SGLT2is, GLP1-RAs, DPP4is, and SUs for Cardiovascular Outcomes.

Abbreviations: CCAE, IBM MarketScan® Commercial Claims and Encounters Data; CVD, cardiovascular disease; GDA, Germany Disease Analyser; IMRD, UK-IQVIA Medical Research Data; MACE, major adverse cardiovascular events; MDCD, IBM Health MarketScan® Multi-State Medicaid Database; MDCR, IBM Health MarketScan® Medicare Supplemental and
Coordination of Benefits Database; OCEDM, Optum® Clinformatics Extended Data Mart - Date of Death; OEHR, Optum® de-identified Electronic Health Record Dataset; OMOP-CDM, Observational Medical Outcomes Partnership-common data model; SIDIAP, Information System for Research in Primary Care; T2DM, type 2 diabetes mellitus; USOC, United States Open Claims; VA, Department of Veterans Affairs Healthcare System.
### Abbreviations:
- **CCAE**: IBM MarketScan® Commercial Claims and Encounters Data
- **DPP4i**: dipeptidyl peptidase 4 inhibitor
- **GDA**: Germany Disease Analyzer
- **GLP1-RA**: glucagon-like peptide-1 receptor agonist
- **IMRD**: UK-IQVIA Medical Research Data
- **MDCD**: IBM Health MarketScan® Multi-State Medicaid Database
- **MDCR**: IBM Health MarketScan® Medicare Supplemental and Coordination of Benefits Database
- **OCEDM**: Optum© Clinformatics Extended Data Mart - Date of Death
- **OEHR**: Optum© de-identified Electronic Health Record Dataset
- **IQR**: interquartile range
- **MDCR**: IBM Health MarketScan® Medicare Supplemental and Coordination of Benefits Database
- **OCEDM**: Optum© Clinformatics Extended Data Mart - Date of Death
- **OEHR**: Optum© de-identified Electronic Health Record Dataset
- **PS**: propensity score
- **SGLT2i**: sodium-glucose co-transporter-2 inhibitor
- **SIDIAP**: Information System for Research in Primary Care
- **SMD**: standardized mean difference
- **SU**: Sulfonylurea
- **USOC**: United States Open Claims
- **VA**: Department of Veterans Affairs Healthcare System
Figure 3: Meta-analytic Calibrated Hazard Ratio Estimates for Comparative Effectiveness of SGLT2is, GLP1-RAs, DPP4is, and SUs for Cardiovascular Outcomes.

**Abbreviations:** DPP4i, dipeptidyl peptidase 4 inhibitor; GLP1-RA, glucagon-like peptide-1 receptor agonist; HF, heart failure; HR, hazard ratio; MACE, major adverse cardiovascular events; MI, myocardial infarction; SGLT2i, sodium-glucose co-transporter-2 inhibitor; SU, Sulfonylurea.
Figure 4: Swarm Plot of Calibrated Hazard Ratio Estimates of Major Outcome Meta-analysis and Leave-one-out Meta-analysis. Circles depict the calibrated relative risk of each leave-one-out study in which one original data source is removed from the meta-analysis. Diamonds depict the original meta-analysis with all sources. Points are color-coded by major outcomes, y-axis includes the medication comparators, and x-axis on log-scale measures the calibrated relative risk of each outcome-comparator combination.

Abbreviations: DPP4i, dipeptidyl peptidase 4 inhibitor; GLP1-RA, glucagon-like peptide-1 receptor agonist; HF, heart failure; MACE, major adverse cardiovascular events; MI, myocardial infarction; SGLT2i, sodium-glucose co-transporter-2 inhibitor; SU, Sulfonylurea.