Circulation

ORIGINAL RESEARCH ARTICLE

Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies

First Results From the RCT DUPLICATE Initiative

BACKGROUND: Regulators are evaluating the use of noninterventional real-world evidence (RWE) studies to assess the effectiveness of medical products. The RCT DUPLICATE initiative (Randomized, Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology) uses a structured process to design RWE studies emulating randomized, controlled trials (RCTs) and compare results. We report findings of the first 10 trial emulations, evaluating cardiovascular outcomes of antidiabetic or antiplatelet medications.

METHODS: We selected 3 active-controlled and 7 placebo-controlled RCTs for replication. Using patient-level claims data from US commercial and Medicare payers, we implemented inclusion and exclusion criteria, selected primary end points, and comparator populations to emulate those of each corresponding RCT. Within the trial-mimicking populations, we conducted propensity score matching to control for >120 preexposure confounders. All study measures were prospectively defined and protocols registered before hazard ratios and 95% CIs were computed. Success criteria for the primary analysis were prespecified for each replication.

RESULTS: Despite attempts to emulate RCT design as closely as possible, differences between the RCT and corresponding RWE study populations remained. The regulatory conclusions were equivalent in 6 of 10. The RWE emulations achieved a hazard ratio estimate that was within the 95% CI from the corresponding RCT in 8 of 10 studies. In 9 of 10, either the regulatory or estimate agreement success criteria were fulfilled. The largest differences in effect estimates were found for RCTs where second-generation sulfonylureas were used as a proxy for placebo regarding cardiovascular effects. Nine of 10 replications had a standardized difference between effect estimates of <2, which suggests differences within expected random variation.

CONCLUSIONS: Agreement between RCT and RWE findings varies depending on which agreement metric is used. Interim findings indicate that selection of active comparator therapies with similar indications and use patterns enhances the validity of RWE. Even in the context of active comparators, concordance between RCT and RWE findings is not guaranteed, partially because trials are not emulated exactly. More trial emulations are needed to understand how often and in what contexts RWE findings match RCTs.

REGISTRATION: URL: https://www.clinicaltrials.gov; Unique identifiers: NCT03936049, NCT04215523, NCT04215536, NCT03936010, NCT03936036, NCT03936062, NCT03936023, NCT03648424, NCT04237935, NCT04237922.

Jessica M. Franklin, PhD
Elisabetta Patorno, MD, DrPH
Rishi J. Desai, MS, PhD
Robert J. Glynn, PhD, ScD
David Martin, MD, MPH
Kenneth Quinto, MD, MPH
Ajinkya Pawar, PhD
Lily G. Bessette, BS
Hemin Lee, MD, MPH
Elizabeth M. Garry, PhD
Nileesa Gautam, BS
Sebastian Schneeweiss, MD, ScD

Key Words: bias ■ diabetes mellitus ■ dipeptidyl-peptidase IV inhibitors ■ randomized controlled trial ■ sodium-glucose transporter 2 inhibitors

Sources of Funding, see page 1012

https://www.ahajournals.org/journal/circ

Clinical Perspective

What Is New?

- RCT DUPLICATE (Randomized, Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology) aims to systematically calibrate nonrandomized real-world evidence (RWE) against randomized, controlled trial (RCT) evidence; in 10 prospectively planned cardiometabolic RCT emulations, RWE studies agreed with RCT findings when design and analysis principles were met and suitable comparators were chosen.
- Whereas insurance claims data with proper design and analysis were fit for the purpose of estimating treatment effects by emulating these RCTs, other RCTs may require alternative data sources.

What Are the Clinical Implications?

- With data that are fit for purpose and proper design and analysis, causal treatment effects can be estimated through both randomized trials and nonrandomized real-world evidence studies.
- These initial findings of the RCT DUPLICATE program indicate circumstances when RWE may offer causal insights where RCT data are either not available or cannot be quickly or feasibly generated. The goal is to develop a resource of high-quality case studies that demonstrate when RWE studies have come to causal conclusions that may serve as reference points to increase confidence in RWE for decision-making.

Regulators of medical products are taking a fresh look at the potential value of real-world evidence (RWE) for decision-making.[1,2] RWE is the clinical evidence about the potential benefits or harms of medical products derived from the analysis of real-world data (RWD), data relating to patient health status and delivery of health care routinely collected from a variety of sources.[3,4] RWE can rely on either randomized or nonrandomized study designs, but concerns remain about whether nonrandomized RWE can accurately assess the effectiveness of drugs.[5,6] These concerns have been highlighted by the rapid execution and dissemination of a large volume of nonrandomized assessments of treatments for coronavirus disease 2019 (COVID-19) with highly variable quality.[7,8]

Calibration of RWE studies against a known treatment effect is one way to evaluate whether RWE can support causal conclusions in select circumstances if conducted using robust methodology. Several systematic reviews have compared the findings of published noninterventional studies with randomized, controlled trial (RCT) findings,[9–15] but they provided limited insights because they identified a wide variety of trials and compared them against published noninterventional studies that often differed substantially in terms of targeted populations, outcomes, or treatment strategies. However, if RWE studies were designed to mimic corresponding RCTs as closely as possible and used causal study designs and analysis methods,[16] such systematic replication efforts would be helpful to understand whether and under which circumstances one would predictably come to the same conclusions.

Ideally, one would compare findings from both RCTs and RWE against the true benefits and harms of a medical product. In the absence of perfect knowledge, RCTs are widely accepted as the best proxies of the true intended drug effects, with the understanding that not all RCTs are perfect. RCTs submitted for regulatory decisions may on average be of higher quality and appropriate as gold standards but are still subject to sampling variability and potential biases from nonadherence and informative dropout. Any evaluation of the agreement between RWE and RCTs will need to consider this uncertainty.[17]

Identifying the magnitude and direction of residual bias attributable to the nonrandomized study design is the key objective of calibrating RWE against RCTs and imperfect emulation of other design features related to limitations of the available RWD is a nuisance that needs to be minimized. Important risks to emulation success are differences between RWE and RCT in the study population, treatment pattern, outcome measurement, or motivations for patients to adhere to study medications. Even if an RWE study is designed to match the corresponding RCT as closely as possible, emulation of all study components may be impossible.[18]

We launched the RCT DUPLICATE initiative (Randomized, Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology) to compare the findings of RCTs relevant to regulatory decision-making with the findings of noninterventional RWE that emulate the trial design as closely as possible in a consistent, transparent, and reproducible process that would be acceptable to regulators.[19] The goal is 3-fold: (1) to identify a process of transparent RWE development that predefines and preregisters all study measures for a single primary analysis; (2) following this process, to quantify how often RWE studies would come to the same conclusion; and (3) to identify the factors that influence whether these 2 study approaches yield similar results.[20] We report the findings of the first 10 attempted replications of RCTs of antidiabetic medications and antiplatelets.

METHODS

Data used in this study may be licensed from the individual data vendors by qualified researchers trained in human subject confidentiality protocols but cannot be shared by study

authors. Protocols are available for each RWE emulation at ClinicalTrials.gov.

Selected Trials

The process for selecting trials to target for replication, as well as details on the RWE study implementation process, have been described previously.[20] Briefly, we sought to identify published and ongoing RCTs that were relevant to regulatory decision-making and potentially replicable in our RWD sources. We considered a trial to be potentially replicable if we could satisfactorily emulate critical aspects of the trial protocol, including the primary outcome, treatment strategies, and inclusion/exclusion criteria, making minor exceptions for features that we could not emulate exactly given the differences in data sources. The trials were required to be sufficiently well-powered and used in a regulatory context. We did not intend to produce a random sample of trials, but instead sought to create a select group of trials that could likely be emulated in longitudinal claims data. The selection process was driven by the availability of relevant data of sufficient quality as described elsewhere.[21]

We present the emulation of 8 cardiovascular outcome trials of antidiabetic medications and 2 trials of antiplatelets. The antidiabetic trials included 7 published trials that compared addition of a single antidiabetic treatment versus addition of placebo to usual care. Among these 7 were 1 trial of liraglutide, a GLP-1 (glucagon-like peptide 1) receptor agonist (LEADER [Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results])[22]; 3 trials of SGLT-2 (sodium glucose co-transporter 2) inhibitors, including dapagliflozin, empagliflozin, and canagliflozin (DECLARE-TIMI 58 [Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events], EMPA-REG OUTCOME [BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients], and CANVAS [Canagliflozin Cardiovascular Assessment Study])[23–25]; and 3 trials of DPP-4 (dipeptidyl peptidase-4) inhibitors, including linagliptin, sitagliptin, and saxagliptin (CARMELINA [Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus], TECOS [Sitagliptin Cardiovascular Outcomes Study (MK-0431-082)], and SAVOR-TIMI 53 [Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications]).[26–28] We also included results from the emulation of the ongoing CAROLINA (Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes), which compared linagliptin with glimepiride, a second-generation sulfonylurea.[29] In this case, the RWE study was completed and submitted for publication 6 months before the CAROLINA findings were released.[30] The 2 antiplatelet trials compared ticagrelor with clopidogrel (PLATO [Platelet Inhibition and Patient Outcomes])[31] or prasugrel with clopidogrel (TRITON-TIMI 38 [Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38]).[32] All 10 trials considered 3-point major adverse cardiovascular events (MACEs), including a composite of cardiovascular death, myocardial infarction, and stroke, to be the primary end point. TECOS also included hospitalization for unstable angina in the primary MACE definition. DECLARE-TIMI 58 had a coprimary end point of cardiovascular death or hospitalization for heart failure. Five additional antiplatelet trials were identified for replication and were dropped because the emulations were underpowered after all trial exclusions were applied or because they assessed treatments given during hospitalization and could therefore not be emulated with the outpatient dispensing data available in our RWD sources (Table I in the Data Supplement).

Data Sources and Implementation Process

We had 3 US health care claims data sources available for emulation of RCTs: Optum Clinformatics (January 1, 2004 through March 31, 2019), IBM MarketScan (2003 through 2017), and a subset of Medicare Parts A, B, and D (2011 through 2017), including all patients with a diabetes diagnosis. Thus, Medicare data were not used for emulation of the antiplatelet trials. Prediction of the CAROLINA findings was limited to data available at the time of study implementation, through September 2015.[30] Data sources contain deidentified information for all covered health care encounters by patients enrolled in participating health insurance plans, including demographics (age and sex), enrollment start and end dates, dispensed medications with dates, dose, and days supply, and performed procedures and medical diagnoses with an associated service date and setting. Medicare claims capture all deaths administratively, but out-of-hospital deaths are less complete in commercial data sources. Cause of death was not available in all data sources.

We designed our RWE study implementation process to make the design and analysis of the RCT replications as structured, transparent, and reproducible as possible.[20] For each RCT, we began by drafting a protocol for the design and analysis of the RWE emulation in health care claims. A similar protocol template was used for all replications, but specific design elements and operational definitions were chosen on the basis of knowledge of the trial and the likely sources of confounding. Creation of the cohort and study variables was implemented using the Aetion Evidence Platform®,[33,34] which records all contact with the claims data and provides an audit trail recording what analyses were conducted and when. Before finalizing the protocol and proceeding with analysis of study outcomes, we evaluated feasibility and validity, including the covariate balance between study groups and an estimate of statistical power. After these checks, we finalized each protocol, including detailed specification of primary analyses, and registered the protocol. No treatment-specific outcome analyses were conducted until after the final RWE study protocol was fully specified and registered.[35] This process was designed to mimic a regulatory submission process and to ensure that the specific design and analysis choices were not influenced by the RWE study results.[19] Complete time-stamped analysis logs are available for review and regulators are able to reproduce and robustness test the RWE studies through the Aetion Evidence Platform®.[20]

Study Design

In each of 3 databases and for each of 10 trials, we identified new users of the exposure of interest and comparator

drugs from pharmacy claims,[33] beginning at the approval date for the exposure (or later if the approval date was before the beginning of available data) and continuing through the end of available data. For the 7 placebo-controlled trials, we selected active comparator groups as a proxy for placebo, because it is well-known that nonuser comparator groups, including untreated patients with diabetes, can differ substantially from actively treated patients in ways that are poorly captured in claims data.[36,37] Specifically, we used DPP-4 inhibitors as the comparator group in the studies of GLP-1 receptor agonists and SGLT-2 inhibitors, and we used second-generation sulfonylureas as the comparator group in the studies of DPP-4 inhibitors. DPP-4 inhibitors and second-generation sulfonylureas were selected as proxies for placebo because they are antidiabetic treatments that have similar indications to the treatments under study, but they are not known to have any effect on the cardiovascular outcomes of interest based on recent evidence.[26–28] Patients were required to have continuous enrollment in the database for 6 months before initiation of the exposure or comparator treatments, and other inclusion/exclusion criteria adapted from each of the 10 trials were implemented. Because the RWE studies were completely prespecified applying the RCT inclusion and exclusion criteria, we abstained from modeling the RWE population characteristics after the actual trial population.[1] Details of each RWE trial emulation, including CONSORT diagrams for cohort formation, are available in the registered protocols on ClinicalTrials.gov (URL: https://www.clinicaltrials.gov; Unique identifiers: NCT03936049, NCT04215523, NCT04215536, NCT03936010, NCT03936036, NCT03936062, NCT03936023, NCT03648424, NCT04237935, and NCT04237922; link provided in the Supplemental Materials in the Data Supplement).

Statistical Analysis

Within these cohorts, we implemented 1:1 propensity score (PS) nearest-neighbor matching[38] with a caliper of 0.01 on the PS scale to control for >120 potential confounders selected a priori, which were measured during the 6 months before drug initiation. Although the trials being emulated generally had fewer patient characteristics listed, a larger set of covariates is necessary in a nonrandomized study to balance as many potential confounders or confounder proxies as possible and emulate baseline randomization. Covariates included demographics, calendar time of treatment initiation, comorbidities, and relevant disease-specific variables, such as use of cardiovascular and other medications, cardiovascular procedures, and indicators of health care utilization as proxy for overall disease state, care intensity, and surveillance. Because laboratory test results were available only for a subset of the patients in the Optum and MarketScan databases, we did not include them in the PS model, but we evaluated postmatching balance in test results between exposure groups.

The primary outcome for all trial emulations except for DECLARE-TIMI 58 was MACEs, adapted from the definition used in the corresponding trial. When emulating DECLARE-TIMI 58, we did not have sufficient statistical power to proceed with analysis of MACEs; therefore, we analyzed only the coprimary composite end point of hospitalization for heart failure and cardiovascular death. For all trials, we used all-cause death as a proxy for cardiovascular death under the assumption that in these populations, which excluded patients with cancer and many other chronic conditions, the majority of deaths would be attributable to cardiovascular conditions. In each trial, we also selected "tracer" outcomes to evaluate as secondary outcomes to better understand the potential role of residual confounding in explaining any differences observed between the RCT and RWE findings. Tracer outcomes were those with known associations with the drugs under study, either null or non-null. Some control outcomes were secondary end points in the trials being emulated; others were selected based on established knowledge. We estimated hazard ratios (HRs) associated with tracer outcomes in the same PS-matched populations identified for analysis of the primary outcomes.

Follow-up for all outcomes started on the day after treatment initiation and continued in an "on-treatment" approach until treatment discontinuation plus a 30-day grace period, switch to a comparator, occurrence of an event of interest, nursing home admission, insurance disenrollment, or end of the study period, whichever came first. The on-treatment analysis in the RWE study attempted to replicate an intention-to-treat estimate from the RCT with very high treatment compliance.[1,4] HRs and 95% CIs were estimated in the PS-matched cohort using Cox regression. Analyses were conducted in each data source separately and then pooled using a fixed-effects meta-analysis, which was selected owing to the very small number of estimates to pool and the use of a uniform study design across the 3 databases. Prespecified sensitivity analyses included an "as-started" analysis, where patients were not censored for treatment changes but were censored at 365 days of follow-up. Several sensitivity and subgroup analyses were conducted after evaluating results of the prespecified analyses. The study was approved by the Brigham and Women's Hospital institutional review board.

RCT–RWE Agreement Assessment

The primary objective of this study was to assess the magnitude of and reasons for differences between RCT findings and findings from corresponding RWE emulations.[20] We prespecified 3 binary agreement metrics: (1) "regulatory agreement" was defined as the ability of the RWE study to replicate the direction and statistical significance of the RCT finding; (2) "estimate agreement" was defined as a RWE HR estimate that was within the 95% CI for the RCT estimate; (3) we conducted hypothesis tests to evaluate whether there was a difference in findings by calculating the standardized difference between the RCT and RWE effect estimates.[20] According to regulatory convention, we consider a P value < 0.05 statistically significant.

Comparator emulation was considered good if the RCT had an active comparator; moderate if a placebo comparison was emulated by an alternative drug thought to be unrelated with the end point of interest, and it was shown to be used in patients with highly similar characteristics, as shown in the covariate balance; and poor if a placebo comparison was emulated by an alternative drug thought to be unrelated to the end point of interest, but it was shown to be used in patients with different characteristics, as shown in the covariate balance. End point emulation was considered

good if the trial outcome could be assessed with high specificity, and moderate if key aspects of the RCT outcome definition were likely to be captured with lower specificity, as shown in the event rates.

Analyses for 5 trials (LEADER, CANVAS, CARMELINA, TECOS, and SAVOR-TIMI 53) were conducted first in early 2019 using data available at that time. These trial emulations were rerun, using identical protocols, in early 2020, as documented in the posted protocols on ClinicalTrials.gov. The reanalyses were conducted to incorporate updated data and to allow for all emulations of the 9 published trials in this report to use the databases over an identical time period. Results from the earlier analyses are available in Table II in the Data Supplement.

RESULTS

Patient Characteristics

Mean or median age across all 10 trials ranged from 61 to 66 years (Table 1). The mean age in each emulation was generally similar to the mean in the corresponding trial, except for emulations of DPP-4 inhibitor trials (CARMELINA, TECOS, SAVOR-TIMI 53, and CAROLINA), which resulted in populations that were slightly older than the corresponding trials despite the same age inclusion criteria. All RWE emulations except TRITON also contained a higher proportion of women than the RCTs. Rates of measured cardiovascular risk factors were generally similar between the RCTs and corresponding RWE emulations, including smoking and hypertension; however, patients in the RWE emulations were more likely to have congestive heart failure. Patients in the emulations of antiplatelet trials were also more likely to have diabetes. Good postmatching balance was achieved on all covariates evaluated in the RWE studies, including laboratory test results that were not included in the PS models (see registered protocols).

Event Rates

For all RCTs that used 3-point MACEs as the primary end point, event rates were lower in the RWE emulation. For example, the event rates in LEADER were 3.4 and 3.9 per 100 person-years in the exposure and comparator groups, versus the emulation event rates of 2.0 and 2.8 (Table 2). However, the 2 trials that had other end points targeted by emulation had higher event rates in the emulations. Specifically, DECLARE-TIMI58 had rates of hospitalization for heart failure or cardiovascular death of 1.2 and 1.5 per 100 person-years in the exposure and comparator groups, versus the emulation event rates of 1.6 and 2.4. TECOS had rates of 3-point MACEs plus hospitalization for acute angina of 4.1 and 4.2 per 100 person-years in the exposure and comparator groups, respectively, versus the emulation event rates of 7.3 and 8.3. These differences in event rates may be attributable to differences in study populations, but concerns about lower specificity of event capture in the RWE led us to label end point emulation as moderate for these trials.

The trends in event rates between the RCTs and corresponding emulations were also reflected in Kaplan-Meier plots (Figure 1). The numbers of patients remaining in the RWE emulations declined quickly in the first 6 months of follow-up, leading to shorter average follow-up in the RWE studies. Event counts in the RWE studies were generally accumulated through a larger number of patients; the RCTs had fewer patients, but longer follow-up. Across all trial emulations, a majority of patients were censored because of discontinuation of their index exposure (Table III in the Data Supplement).

RCT–RWE Agreement

Regulatory agreement was found for 6 of 10 emulations (Figure 2). PLATO found ticagrelor to be superior to clopidogrel (HR, 0.84 [95% CI, 0.77, 0.92]); the emulation found an HR point estimate in the same direction, but with an upper 95% CI limit above 1.0 (HR, 0.92 [95% CI, 0.83, 1.02]). The 3 RCTs of DPP-4 inhibitors versus placebo found them to be noninferior but not superior to placebo with respect to cardiovascular risk (CARMELINA, TECOS, SAVOR-TIMI 53). The emulations of those trials also found noninferiority, but they additionally found superiority. Estimate agreement was achieved for all trials except for DECLARE and SAVOR-TIMI 53, where the emulation estimates were below the lower 95% CI bound from the RCTs. SAVOR-TIMI 53 was the only trial to have a statistically significant difference between the trial and emulation estimate (standardized difference, 3.16).

Database-specific estimates did not indicate that a single database was consistently leading to higher or lower effect estimates and variation across databases was in line with what would be expected given CI width (Figure I in the Data Supplement). Sensitivity analyses did not produce meaningful changes in study findings (Table IV in the Data Supplement). "As started" analyses generally resulted in estimated HRs closer to null, likely because of increased exposure misclassification over time in those analyses. Other sensitivity analyses did not produce consistent shifts in estimated treatment effects across trial emulations.

Analysis of tracer outcomes in RWE generally returned expected findings, indicating low potential for residual confounding related to these outcomes (Table 3). The only exception was in the analysis of pneumonia hospitalization in the emulation of TRITION, which estimated a decreased risk for ticagrelor versus clopidogrel, despite no expected effect on this outcome. The risk-adjustment strategy from the main study did not focus on predictors of pneumonia.

Table 1. Patient Characteristics in the Randomized, Controlled Trials (RCTs) Versus the Corresponding Real-World Evidence (RWE) Emulations

Characteristics	LEADER	DECLARE-TIMI58	EMPA-REG OUTCOME	CANVAS	CARME-LINA	TECOS	SAVOR-TIMI 53	CARO-LINA	TRITON-TIMI 38	PLATO
Age, y, mean±SD										
RCT	64.3±7.2	64.0±6.8	63.1±8.7	63.3±8.3	65.9±9.1	65.5±8.0	65.1±8.6	64.1±9.5	61*	62*
RWE	67.7±6.0	62.6±5.9	61.9±9.6	65.3±6.6	72.3±9.0	72.3±8.6	68.9±7.8	70.4±7.7	56.9±10.0	65.4±10.8
Female, %										
RCT	35.7	37.4	28.6	35.8	37.1	29.3	33.1	40.0	26.0	28.4
RWE	53.5	42.3	40.5	46.3	53.3	47.4	46.8	52.3	21.1	32.5
Smoking, %										
RCT	12.1	14.5	13.2	17.8	10.2	11.4	—	19.7	38.0	35.9
RWE	10.2	7.8	9.4	8.4	11.5	15.7	7.8	6.0	36.7	37.4
History of MI, %										
RCT	30.7	—	46.6	—	—	42.6	37.8	—	18.0	20.5
RWE	13.0	9.3	10.3	8.8	21.3	29.9	11.2	9.0	17.2	23.7
Hypertension, %										
RCT	90.0	89.4	95.0†	90.0	91.0	≥78.8†	81.8	90.1	64.0	65.4
RWE	97.1	88.3	90.9	96.6	96.4	94.0	88.7	92.7	51.1	70.2
CHF, %										
RCT	17.9	10.1	10.1	14.4	26.8	18.0	12.8	4.5	—	5.6
RWE	20.9	11.4	11.8	12.6	33.2	39.3	17.6	12.1	5.0	11.9
Diabetes, %										
RCT	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	23.0	25.0
RWE	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	28.1	47.4
Previous CVD, %										
RCT	81.4	40.6	75.6	65.6	58.1	74.0	78.6	41.9	—	—
RWE	63.0	51.4	45.9	45.4	88.1	100.0	56.3	54.4	37.2	54.7

CANVAS indicates Canagliflozin Cardiovascular Assessment Study; CARMELINA, Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus; CAROLINA, Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes; CHF, congestive heart failure; CVD, cardiovascular disease; DECLARE-TIMI58, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; MI, myocardial infarction; PLATO, Platelet Inhibition and Patient Outcomes; SAVOR-TIMI 53, Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications; TECOS, Sitagliptin Cardiovascular Outcomes Study (MK-0431-082); and TRITON-TIMI 38, Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38.

*Reported median age only.

†Rates of hypertension assumed from reported use of antihypertensive medications.

DISCUSSION

RCT DUPLICATE seeks to provide rigorously derived evidence for a selected sample of trials and end points on when and how noninterventional RWE studies reach the same conclusions as RCTs that were conducted in a regulatory context.[20] In this interim report on the first 10 completed emulations of RCTs, we found that 6 out of 10 emulations met the criteria for full regulatory agreement. Eight out of 10 emulations achieved estimate agreement. In only 1 emulation, the standardized difference was >2 (P=0.002).

Some emulations would be expected to fail to produce findings similar to the RCT, even in the absence of any bias,[20] just as RCTs sometimes fail to replicate previous RCT findings.[17] The probability of regulatory agreement in the absence of bias was estimated to be in the 80% to 90% range for RCTs that found statistically significant effects and potentially much lower for RCTs that failed to find significant effects. When the variances of the 2 estimates are equal, which they nearly were in most cases here, there is an 83% chance of estimate agreement in the absence of any bias in the RWE studies, and we would expect 5% of emulations to have a standardized difference >2. In our report, 1 of 10 emulations had a standardized difference of >2 (ie, P value <0.05). However, it is also possible that an emulation could result in agreement with the corresponding RCT attributable solely to random variation, despite a large systematic bias in the design.

Overall, agreement between the RCT and RWE estimates was good for all antidiabetic trials except those that compared a DPP-4 inhibitor with placebo. In emulations of these trials, second-generation sulfonylureas

Table 2. Study Sizes and Event Rates

		RCT						RWE					
		Exposure			Comparator			Exposure			Comparator		
Study	Outcome	Events	N	Rate*	Events	N	Rate*	Events	N	Rate*	Events	N	Rate*
LEADER	3P MACE	608	4668	3.4	694	4672	3.9	1352	84346	2.1	1955	84346	2.6
DECLARE-TIMI 58	HHF+cardiovascular death	417	8582	1.2	496	8578	1.5	242	24895	1.6	367	24895	2.4
EMPA-REG OUTCOME	3P MACE	490	4687	3.7	282	2333	4.4	416	51875	1.5	478	51875	1.9
CANVAS	3P MACE	564†	5795	2.7	496†	4347	3.2	772	76099	1.5	990	76099	1.9
CARMELINA	3P MACE	434	3494	5.8	420	3485	5.6	1540	50913	4.6	1826	50913	5.2
TECOS	3P MACE+angina	839	7257	4.1	851	7266	4.2	8106	174739	7.3	9692	174739	8.3
SAVOR-TIMI 53	3P MACE	613	8280	3.6	609	8212	3.6	1662	91064	2.4	2390	91064	3.1
CAROLINA	3P MACE	356	3023	2.1	362	3010	2.1	373	24131	2.7	458	24131	3.0
TRITON-TIMI 38	3P MACE	643	6813	7.9	781	6795	9.7	718	21932	3.8	960	21932	3.9
PLATO	3P MACE	864	9333	9.8	1014	9291	11.7	649	13980	8.0	858	13980	7.1

3P MACE indicates 3-point major adverse cardiovascular events (myocardial infarction, stroke, or cardiovascular death); CANVAS, Canagliflozin Cardiovascular Assessment Study; CARMELINA, Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus; CAROLINA, Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes; CHF, congestive heart failure; CVD, cardiovascular disease; DECLARE-TIMI 58, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; HHF, hospitalization for heart failure; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; PLATO, Platelet Inhibition and Patient Outcomes; RWE, real-world evidence; SAVOR-TIMI 53, Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications; TECOS, Sitagliptin Cardiovascular Outcomes Study (MK-0431-082); and TRITON-TIMI 38, Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38.

*Incidence rate per 100 person-years.

†Not reported in the randomized, controlled trial. Estimated on the basis of reported event rate and mean follow-up time.

were used as a proxy for placebo. Active comparators can decrease confounding if they are used interchangeably but they are still not a perfect emulation of a placebo add-on group. If the older and less expensive sulfonylureas were used more often in patients with unmeasured frailty and lower socioeconomic status, then bias toward a protective effect for DPP-4 inhibitors would be expected, as was found in the emulations. In contrast, all other diabetes trial emulations targeted a trial with an active comparator or used DPP-4 inhibitors as a proxy for placebo. These findings reinforce the widespread recommendations to select active comparators that are used in similar indicated populations in noninterventional studies using health care databases.[21,35,39]

The 2 antiplatelet trials did not include placebo, so selection of active comparators did not present a challenge. Regulatory and estimate agreement was observed in TRITON-TIMI 38 and estimate agreement was observed in PLATO. Sensitivity analyses and analysis of tracer outcomes did not reveal a specific design-related hypothesis for the lack of a superiority finding in the RWE emulation of PLATO. Additional hypotheses include random error and a lack of Medicare data for this therapeutic area, which presumably lowered the event rate for these emulations.

Despite substantial effort, this activity, and arguably any project of this type, has important limitations. The emulation of a trial requires many subjective decisions regarding how to emulate the RCT design and how to control confounding. Although we attempted to emulate the features of each targeted RCT as closely as possible, including inclusion and exclusion criteria, exposures, and outcomes, the constraints of the health care databases made exact emulation impossible. Close emulation of placebo is impossible via RWD, and selection of an active placebo proxy may fundamentally change the study question. Adherence to medications used in routine care is often poor compared with RCTs. We attempted to account for poor adherence to study medications in the RWE by conducting on-treatment analyses that censor patients at treatment discontinuation, which led to shorter average follow-up time in the RWE versus RCTs. This may have excluded some important outcome events in the RWE if patients were discontinuing or switching their medication because of poor prognosis. In contrast, the RCTs typically used an intention-to-treat approach for the primary analysis, which is known to result in effect estimates closer to the null in the context of medication nonadherence.[40] This difference may partly explain the larger effects observed in several RWE emulations versus RCTs.

Because complete health history and laboratory data were not available for all patients in the RWE, inclusion and exclusion criteria from the trials could only be partially emulated, and even where fully emulated (eg, age, the resulting distributions were at times meaningfully different between the RCT and RWE populations,

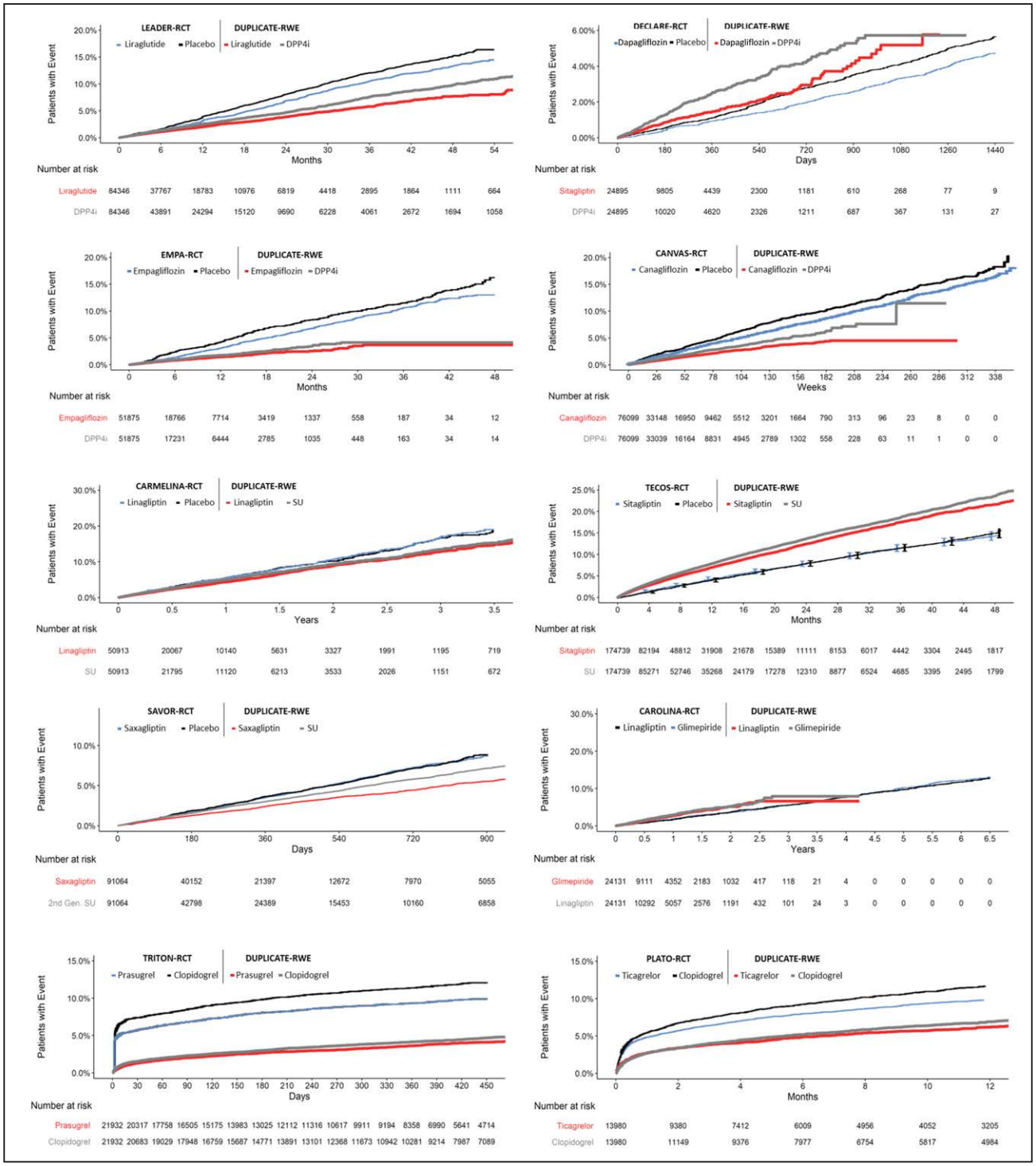

Figure 1. Comparison of cumulative event curves.

Cumulative event Kaplan-Meier plots for primary end points in the randomized, controlled trials (RCTs) and corresponding real-world evidence (RWE) emulations. CANVAS indicates Canagliflozin Cardiovascular Assessment Study; CARMELINA, Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus; CAROLINA, Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes; DECLARE, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events; DPP-4, dipeptidyl peptidase-4; EMPA, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; PLATO, Platelet Inhibition and Patient Outcomes; SAVOR, Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications; TECOS, Sitagliptin Cardiovascular Outcomes Study (MK-0431-082); TRITON, Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38; and SU, sulfonylureas.

possibly because of nonrepresentative participation in RCTs). Cause of death was not available in the claims databases, so cardiovascular death was approximated by counting all recorded deaths, which did not completely capture out-of-hospital deaths for the commercial claims databases used in this study. Such measurement

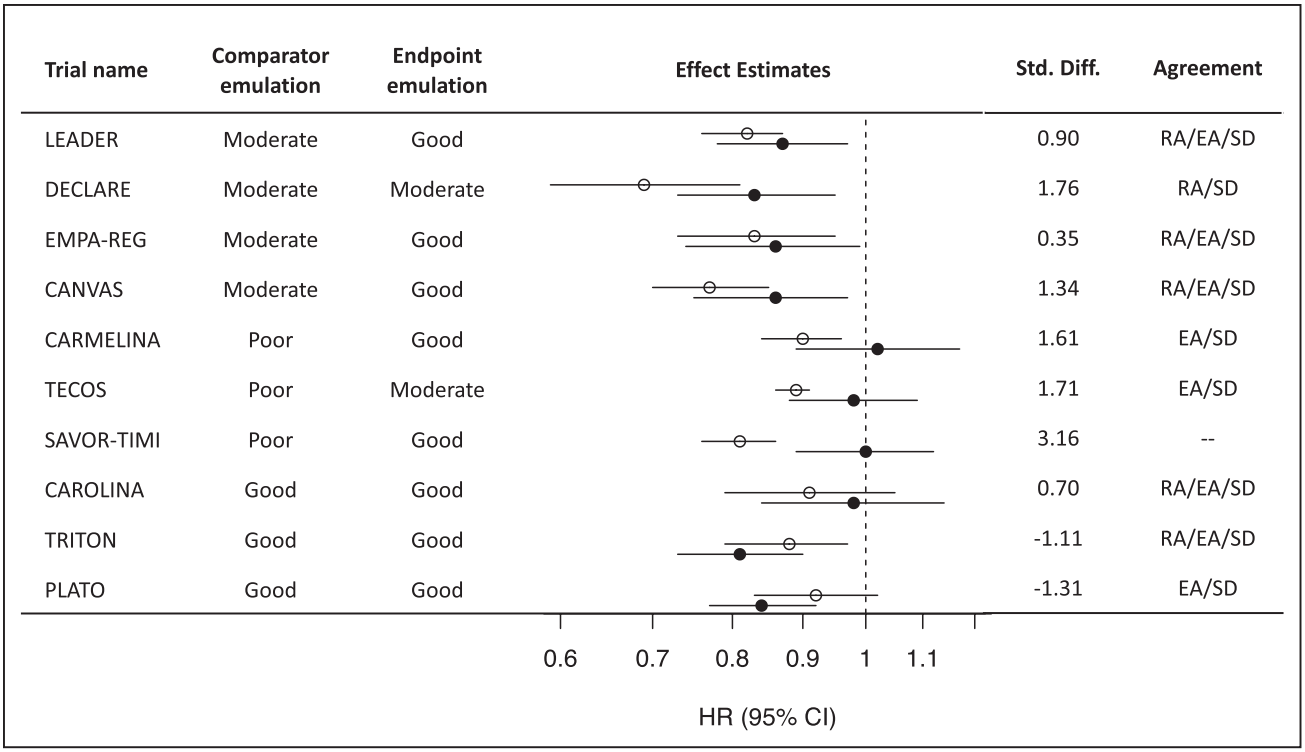

Figure 2. Agreement between randomized, controlled trial (RCT) findings and their prespecified real-world evidence (RWE) emulations.
Open circles represent the estimated hazard ratio (HR) from RWE, and filled circles represent the estimated HR from the corresponding RCT. Under the null hypothesis of no bias in the RWE, we would expect ≈5% of emulations to have a standardized difference >2. CANVAS indicates Canagliflozin Cardiovascular Assessment Study; CARMELINA, Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus; CAROLINA, Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes; DECLARE, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events; EA, estimate agreement reached; EMPA-REG, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; HR, hazard ratio; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; PLATO, Platelet Inhibition and Patient Outcomes; RA, regulatory agreement reached; SAVOR-TIMI, Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications; SD, standardized difference <1.96; TECOS, Sitagliptin Cardiovascular Outcomes Study (MK-0431-082); and TRITON, Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38.

error in the outcome can lead to conservative treatment effect estimates when misclassification is nondifferential across treatment groups.[41] Furthermore, the specificity of all-cause death as a measure of cardiovascular death in patient populations with underlying conditions, such as the diabetes and acute coronary syndrome populations evaluated in this article, would be expected to be better than in the general population.

The use of claims data, which lack clinical detail but provide longitudinal data across the care continuum, affected the agreement between RCT and RWE findings. Other RWD sources, such as electronic health records and patient registries, would almost certainly have led to different results, as they often have detailed clinical information that may improve confounding adjustment. On the other hand, some data elements are better captured in claims data than in electronic health records, which may overestimate medication use by patients who fail to fill their prescriptions and may miss outcomes treated by out-of-system providers, resulting in substantial bias.[42] Other differences in capturing outcome events are also likely. As noted earlier, all of these differences are failures of emulation, but do not represent bias attributable to lack of randomization in the RWE, which is the primary focus of this project.[18]

The level of agreement between RWE and RCTs reported here was achieved with a prespecified RWE protocol. This protocol detailed the primary analytic strategy and was publicly registered before any comparative analyses of outcomes between treatment groups were conducted, as documented in our audit trail in the Aetion Evidence Platform® and viewable by US Food and Drug Administration investigators. These prespecified analyses represent the agreement that could be expected when designing an RWE study without knowledge of the findings of the possibly hypothetical RCT that is the target of emulation.[43] The findings in this report are therefore relevant for the interpretation of nonrandomized studies for clinical decision-making when there is no RCT evidence available on a given question. However, interpretation of findings is limited to the small subset of clinical questions that can support measurement of necessary inclusion and exclusion criteria, exposures, outcomes, and confounders in the claims databases available to us for this project. Interpretation is further limited to nonrandomized studies that base their design on emulation of a potentially hypothetical target trial

Table 3. Effect Estimates for Tracer Outcomes

Study	Outcome	Expected HR*	Exposure IR†	Comparator IR†	Observed HR (95% CI)
LEADER	Severe hypoglycemia	<1	7.8	10.5	0.73 (0.65–0.81)
DECLARE-TIMI58	Diabetic ketoacidosis	>1	2.0	1.4	1.36 (0.78–2.37)
EMPA-REG OUTCOME	HF hospitalization	<1	2.6	7.7	0.35 (0.27–0.46)
	Diabetic ketoacidosis	>1	2.9	2.3	1.25 (0.89–1.76)
CANVAS	HF hospitalization	<1	2.8	7.8	0.36 (0.30–0.44)
	Diabetic ketoacidosis	>1	2.6	1.5	1.70 (1.29–2.25)
CARMELINA	ESRD	≈1	3.2	3.2	1.04 (0.81–1.33)
TECOS	Severe hypoglycemia	<1	12.3	30.8	0.40 (0.38–0.43)
SAVOR-TIMI 53	Severe hypoglycemia	<1	5.9	16.3	0.37 (0.33–0.41)
CAROLINA	Severe hypoglycemia	<1	6.0	16.0	0.42 (0.32–0.56)
	ESRD	≈1	3.0	3.2	1.08 (0.66–1.79)
TRITON-TIMI 38	Major bleeding	>1	20.2	16.0	1.17 (1.01–1.34)
	Pneumonia hospitalization	≈1	11.5	12.3	0.83 (0.73–0.95)
PLATO	Major bleeding	≈1	29.4	23.0	1.16 (0.98–1.39)
	Pneumonia hospitalization	≈1	23.4	22.0	1.01 (0.84–1.22)

CANVAS indicates Canagliflozin Cardiovascular Assessment Study; CARMELINA, Cardiovascular and Renal Microvascular Outcome Study With Linagliptin in Patients With Type 2 Diabetes Mellitus; CAROLINA, Cardiovascular Outcome Study of Linagliptin Versus Glimepiride in Patients With Type 2 Diabetes; DECLARE-TIMI58, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; ESRD, end stage renal disease; HF, heart failure; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; PLATO, Platelet Inhibition and Patient Outcomes; SAVOR-TIMI 53, Does Saxagliptin Reduce the Risk of Cardiovascular Events When Used Alone or Added to Other Diabetes Medications; TECOS, Sitagliptin Cardiovascular Outcomes Study (MK-0431-082); and TRITON-TIMI 38, Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel–Thrombolysis in Myocardial Infarction 38.

*An expected hazard ratio (HR) of ≈1 indicates an approximately null effect. Other expected HRs are listed as ranges of either >1 or <1.

†Incidence rate (IR) per 1000 person-years.

and would not directly generalize to the many nonrandomized RWE studies that do not follow this strategy.

The past few years have witnessed rapid growth in RWE, including an explosion of RWE on treatments for COVID-19,[7,8] increasing use of single-arm studies and nonrandomized studies for US Food and Drug Administration approval and labeling,[44] and the use of nonrandomized studies by payors and health systems to make formulary decisions. Additional growth is expected as access to and reliability of RWD sources continues to mature. Whereas the quality of nonrandomized RWE studies varies, explicit prespecification of a protocol corresponding to a target trial, as demonstrated in our study, could increase confidence when results of such studies are used for regulatory and payor decisions. In addition, sharing of data and analyses resulting from large health care databases with regulators could strengthen the credibility of nonrandomized RWE. Although such data are typically protected by patient privacy regulations and are licensed to investigators by data vendors that prohibit the sharing of patient-level data, access to data may be provided through virtual data enclaves or analytics platforms, such as the one used in this research.

Conclusions

Agreement between RCT and RWE findings varies depending on which agreement metric is used. Interim findings confirm that selection of active comparator therapies with similar indications and use patterns enhances the validity of RWE. However, even in the context of active comparators, concordance between RCT and RWE findings is not guaranteed. Our findings are based on a select sample of studies where RWD could emulate the outcome measure satisfactorily and key confounding factors were observable. More evidence is needed to contribute to understanding of the circumstances that determine whether RWE findings can predictably match those of RCTs across different therapeutic areas. The RCT DUPLICATE project will continue and expand to include several therapeutic areas over the course of the project.

ARTICLE INFORMATION

Received October 6, 2020; accepted November 30, 2020.

The Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/circulationaha.120.051718.

Correspondence

Jessica M. Franklin, PhD, 1620 Tremont St, Suite 3030, Boston, MA 02120. Email JMFranklin@bwh.harvard.edu

Affiliations

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA (J.M.F., E.P., R.J.D., R.J.G., A.P., L.G.B., H.L., N.G., S.S.). Office of Medical Policy, Center for Drug Evaluation and Research, US Food and Drug

Administration, Silver Spring, MD (D.M., K.Q.). Scientific Research, Aetion Inc, Boston, MA (E.M.G.).

Acknowledgments

The authors thank their external advisor group for input throughout the project and review of study protocols: Professors Wayne Ray (Vanderbilt University), Miguel Hernan (Harvard University), Samy Suissa (McGill University), Steve Goodman (Stanford University), and Alan Brookhart (Duke University); US Food and Drug Administration colleagues Dianne Paraoan for project management, Mark Levenson and Robert Temple for input on the design of the RCT-DUPLICATE initiative, and members of multiple US Food and Drug Administration review divisions for comments on trial emulations; and Danielle Isaman, Liza Gibbs, and Katherine Gilpin (Aetion Inc) for implementation support using the Aetion Evidence Platform®.

Sources of Funding

This study was funded by contracts from the Food and Drug Administration (contracts HHSF223201710186C and HHSF223201810146C) to the Brigham and Women's Hospital and Aetion Inc. Dr Patorno was supported by career development grant K08AG055670 from the National Institute on Aging.

Disclosures

Dr Schneeweiss is principal investigator of the US Food and Drug Administration Sentinel Innovation Center funded by the US Food and Drug Administration; co–principal investigator of an investigator-initiated grant to the Brigham and Women's Hospital from Boehringer Ingelheim unrelated to the topic of this study; is a consultant to Aetion Inc., a software manufacturer of which he owns equity; his interests were declared, reviewed, and approved by the Brigham and Women's Hospital and Partners HealthCare System in accordance with their institutional compliance policies. Dr Patorno is coinvestigator of an investigator-initiated grant to the Brigham and Women's Hospital from Boehringer-Ingelheim, not directly related to the topic of the submitted work. Dr Desai has served as principal investigator for research grants from Bayer, Vertex, and Novartis to the Brigham and Women's Hospital for unrelated projects. Dr Glynn has received research support from investigator-initiated grants to the Brigham and Women's Hospital for clinical trials funded by AstraZeneca, Kowa, Pfizer, and Novartis. Dr Garry is an employee of Aetion Inc with stock options. The views expressed in the article are the personal views of the authors and may not be understood, quoted, or stated on behalf of or reflecting the views or policies of the Department of Health and Human Services or the US Food and Drug Administration. The other authors report no conflicts.

Supplemental Materials

Supplemental Materials and Methods
Data Supplement Tables I–IV
Data Supplement Figure I
References 45 and 46

REFERENCES

1. Bonamici S. H.R.34–21st Century Cures Act. https://www.congress.gov/bill/114th-congress/house-bill/34. Published 2016. Accessed June 15, 2017.
2. US Food and Drug Administration. PDUFA VI: Fiscal Years 2018–2022. https://www.fda.gov/industry/prescription-drug-user-fee-amendments/pdufa-vi-fiscal-years-2018-2022. Published 2017. Accessed June 15, 2017.
3. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, et al. Real-world evidence: what is it and what can it tell us? *N Engl J Med.* 2016;375:2293–2297. doi: 10.1056/NEJMsb1609216
4. Jarow JP, LaVange L, Woodcock J. Multidimensional evidence generation and FDA regulatory decision making: defining and using "real-world" data. *JAMA.* 2017;318:703–704. doi: 10.1001/jama.2017.9991
5. Slattery J, Kurz X. Assessing strength of evidence for regulatory decision making in licensing: what proof do we need for observational studies of effectiveness? *Pharmacoepidemiol Drug Saf.* 2020;29:1336–1340. doi: 10.1002/pds.5005
6. Schneeweiss S. Real-world evidence of treatment effects: the useful and the misleading. *Clin Pharmacol Ther.* 2019;106:43–44. doi: 10.1002/cpt.1405
7. Pundi K, Perino AC, Harrington RA, Krumholz HM, Turakhia MP. Characteristics and strength of evidence of COVID-19 studies registered on ClinicalTrials.gov. *JAMA Intern Med.* 2020;180:1398–1400. doi: 10.1001/jamainternmed.2020.2904
8. Califf RM, Hernandez AF, Landray M. Weighing the benefits and risks of proliferating observational treatment assessments: observational cacophony, randomized harmony. *JAMA.* 2020;324:625. doi: 10.1001/jama.2020.13319
9. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev.* 2014;4:MR000034. doi: 10.1002/14651858.MR000034.pub2
10. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887–1892. doi: 10.1056/NEJM200006223422507
11. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA.* 2001;286:821–830. doi: 10.1001/jama.286.7.821
12. Bhandari M, Tornetta P III, Ellis T, Audige L, Sprague S, Kuo JC, Swiontkowski MF. Hierarchy of evidence: differences in results between non-randomized studies and randomized trials in patients with femoral neck fractures. *Arch Orthop Trauma Surg.* 2004;124:10–16. doi: 10.1007/s00402-003-0559-z
13. Dahabreh IJ, Sheldrick RC, Paulus JK, Chung M, Varvarigou V, Jafri H, Rassen JA, Trikalinos TA, Kitsios GD. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *Eur Heart J.* 2012;33:1893–1901. doi: 10.1093/eurheartj/ehs114
14. Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, Ravaud P. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg.* 2014;259:18–25. doi: 10.1097/SLA.0000000000000256
15. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ.* 2016;352:i493. doi: 10.1136/bmj.i493
16. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol.* 2016;183:758–764. doi: 10.1093/aje/kwv254
17. Franklin JM, Dejene S, Huybrechts KF, Wang SV, Kulldorff M, Rothman KJ. A bias in the evaluation of bias comparing randomized trials with nonexperimental studies. *Epidemiol Methods.* 2017;6:20160018. doi: 10.1515/em-2016-0018
18. Franklin JM, Glynn RJ, Suissa S, Schneeweiss S. Emulation differences vs. biases when calibrating real-world evidence findings against randomized controlled trials. *Clin Pharmacol Ther.* 2020;107:735–737. doi: 10.1002/cpt.1793
19. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther* 2019;105:867–877. doi: 10.1002/cpt.1351
20. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, Schneeweiss S. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther.* 2020;107:817–826. doi: 10.1002/cpt.1633
21. Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther.* 2017;102:924–933. doi: 10.1002/cpt.857
22. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, Nissen SE, Pocock S, Poulter NR, Ravn LS, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2016;375:311–322. doi: 10.1056/NEJMoa1603827
23. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2019;380:347–357. doi: 10.1056/NEJMoa1812389
24. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, et al; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med.* 2015;373:2117–2128. doi: 10.1056/NEJMoa1504720

25. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, Shaw W, Law G, Desai M, Matthews DR, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med.* 2017;377:644–657. doi: 10.1056/NEJMoa1611925
26. Rosenstock J, Perkovic V, Johansen OE, Cooper ME, Kahn SE, Marx N, Alexander JH, Pencina M, Toto RD, Wanner C, et al. Effect of linagliptin vs placebo on major cardiovascular events in adults with type 2 diabetes and high cardiovascular and renal risk: the CARMELINA randomized clinical trial. *JAMA.* 2019;321:69. doi: 10.1001/jama.2018.18269
27. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, Josse R, Kaufman KD, Koglin J, Korn S, et al; TECOS Study Group. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2015;373:232–242. doi: 10.1056/NEJMoa1501352
28. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, Ohman P, Frederich R, Wiviott SD, Hoffman EB, et al; SAVOR-TIMI 53 Steering Committee and Investigators. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med.* 2013;369:1317–1326. doi: 10.1056/NEJMoa1307684
29. Rosenstock J, Kahn SE, Johansen OE, Zinman B, Espeland MA, Woerle HJ, Pfarr E, Keller A, Mattheus M, Baanstra D, 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. doi: 10.1001/jama.2019.13772
30. Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D, Franklin JM. Using real-world data to predict findings of an ongoing phase iv cardiovascular outcome trial: cardiovascular safety of linagliptin versus glimepiride. *Diabetes Care.* 2019;42:2204–2210. doi: 10.2337/dc19-0069
31. Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horrow J, Husted S, James S, Katus H, et al; PLATO Investigators. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med.* 2009;361:1045–1057. doi: 10.1056/NEJMoa0904327
32. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann FJ, Ardissino D, De Servi S, Murphy SA, et al; TRITON-TIMI 38 Investigators. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med.* 2007;357:2001–2015. doi: 10.1056/NEJMoa0706482
33. Wang SV, Verpillat P, Rassen JA, Patrick A, Garry EM, Bartels DB. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clin Pharmacol Ther.* 2016;99:325–332. doi: 10.1002/cpt.329
34. Kim SC, Solomon DH, Rogers JR, Gale S, Klearman M, Sarsour K, Schneeweiss S. Cardiovascular safety of tocilizumab versus tumor necrosis factor inhibitors in patients with rheumatoid arthritis: a multidatabase cohort study. *Arthritis Rheumatol.* 2017;69:1154–1164. doi: 10.1002/art.40084
35. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, Madigan D, Makady A, Schneeweiss S, Tarricone R, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf.* 2017;26:1033–1039. doi: 10.1002/pds.4297
36. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology.* 2001;12:682–689. doi: 10.1097/00001648-200111000-00017
37. Glynn RJ, Schneeweiss S, Wang PS, Levin R, Avorn J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. *J Clin Epidemiol.* 2006;59:819–828. doi: 10.1016/j.jclinepi.2005.12.012
38. Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf.* 2012;21(Suppl 2):69–80. doi: 10.1002/pds.3263
39. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf.* 2010;19:858–868. doi: 10.1002/pds.1926
40. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials.* 2012;9:48–55. doi: 10.1177/1740774511420743
41. Desai RJ, Levin R, Lin KJ, Patorno E. Bias implications of outcome misclassification in observational studies evaluating association between treatments and all-cause or cardiovascular mortality using administrative claims. *J Am Heart Assoc.* 2020;9:e016906. doi: 10.1161/JAHA.120.016906
42. Lin KJ, Glynn RJ, Singer DE, Murphy SN, Lii J, Schneeweiss S. Out-of-system care and recording of patient characteristics critical for comparative effectiveness research. *Epidemiology.* 2018;29:356–363. doi: 10.1097/EDE.0000000000000794
43. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70–75. doi: 10.1016/j.jclinepi.2016.04.014
44. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open.* 2016;6:e011666. doi: 10.1136/bmjopen-2016-011666
45. Patorno E, Gopalakrishnan C, Franklin JM, Brodovicz KG, Masso-Gonzalez E, Bartels DB, Liu J, Schneeweiss S. Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical variables only observed in electronic health records. *Diabetes Obes Metab.* 2018;20:974–984. doi: 10.1111/dom.13184
46. van Onzenoort HA, Menger FE, Neef C, Verberk WJ, Kroon AA, de Leeuw PW, van der Kuy PH. Participation in a clinical trial enhances adherence and persistence to treatment: a retrospective cohort study. *Hypertension.* 2011;58:573–578. doi: 10.1161/HYPERTENSIONAHA.111.171074