Emulating Six Cardiovascular Outcome Trials of Antidiabetic Drugs in Individuals with Type 2 Diabetes: Analyses Based on Real-World Databases from Korea and Taiwan

Hwa Yeon Ko[1], Philip C. M. Au[2], Michael Chun-Yuan Cheng[3,4], Ching-Lung Cheung[2], Ahhyung Choi[1,5], Miyuki Hsing-Chun Hsieh[3,4], Han Eol Jeong[1], Edward Chia-Cheng Lai[3,4], Brian Meng-Hsun Li[3,4], Kenneth K. C. Man[2,6], Jeremy A. Rassen[7], Daniel Hsiang-Te Tsai[3,4], Shirley V. Wang[8], Ian C. K. Wong[2,6,9], Ju-Young Shin[1,10,*] and Sohee Park[1,9,*]

With the increasing volume of clinical evidence derived from large-scale Asian real-world data (RWD) and the growing interest in its use in regulatory decision-making, it is important to conduct benchmarking exercises that evaluate whether studies using Asian RWD can generate similar conclusions as randomized controlled trials (RCTs). We aimed to assess whether observational studies based on Korea and Taiwan RWD can yield comparable results with trials by emulating six cardiovascular outcome trials (CVOTs) of antidiabetic drugs in individuals with type 2 diabetes (T2D). We emulated six CVOTs using nationwide claims of Korea and Taiwan. An active comparator, new-user design was applied, and observational analogues to the eligibility criteria and outcomes of the corresponding RCT were implemented. Propensity score matching was utilized to balance the treatment groups. Hazard ratios and 95% confidence intervals (CI) were estimated and compared with corresponding RCT estimates. We used three pre-specified binary agreement metrics that have been used in prior benchmarking studies to define concordance in results. Results from each of the six emulations were concordant with the corresponding CVOT on ≥1 binary agreement metric. Five out of six emulations indicated superiority when the corresponding CVOT only demonstrated non-inferiority. Results from emulations were more concordant with Asian-specific results from RCT, with four emulations meeting all agreement metrics. In this binational study using two Asian healthcare claims data, emulations yielded comparable clinical conclusions with the corresponding RCT, increasing the confidence in the validity of RWE studies in patients with T2D using these databases.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ There is growing interest in real-world evidence (RWE) derived from observational studies that utilize routinely collected real-world data (RWD) from diverse populations. Attempts to validate RWD and its methodology for causal inference, by emulating and benchmarking against published landmark trials, have gained prominence. With the increasing volume of clinical evidence derived from large-scale Asian RWD, it is important to conduct benchmarking exercises that evaluate whether studies using Asian RWD can generate similar conclusions as randomized controlled trials.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ Can observational studies using RWD from Asia generate conclusions similar to those of randomized controlled trials (RCTs)?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ This binational study emulated six cardiovascular outcome trials (CVOTs) of antidiabetic drugs in patients with type 2 diabetes (T2D) using Korea and Taiwan RWD. The concordance in results demonstrates the ability of these databases to capture key study parameters relevant to T2D and supports the notion that when coupled with rigorous methods, these databases can generate valid and robust causal conclusions about drug effects within the T2D therapeutic area.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Database studies conducted with RWD from Korea and Taiwan can provide important complementary evidence to RCTs, especially for underrepresented Asian populations in RCTs; thereby supporting decision-making in both clinical and regulatory settings.

Received February 6, 2025; accepted July 28, 2025. doi:10.1002/cpt.70027

[1]School of Pharmacy, Sungkyunkwan University, Suwon, South Korea; [2]Department of Pharmacology and Pharmacy, Centre for Safe Medication Practice and Research, University of Hong Kong, Hong Kong City, Hong Kong; [3]School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [4]Population Health Data Center, National Cheng Kung University, Tainan, Taiwan; [5]Harvard-MIT Center for Regulatory Science, Harvard Medical School, Boston, Massachusetts, USA; [6]Research Department of Practice and Policy, School of Pharmacy, University College London, London, UK; [7]Aetion, New York, New York, USA; [8]Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [9]Aston Pharmacy School, Aston University, Birmingham, UK; [10]Department of Biohealth Regulatory Science, Sungkyunkwan University, Suwon, South Korea. *Correspondence: Ju-Young Shin (shin.jy@skku.edu) and Sohee Park (s.park3@aston.ac.uk)

There is growing interest in real-world evidence (RWE) derived from observational studies that utilize routinely collected real-world data (RWD) from diverse population.[1–4] In alignment with this focus, attempts to validate RWD and its methodology for causal inference, by emulating and benchmarking against published landmark trials, have gained prominence.[5–8] It is necessary to evaluate the relevance and reliability of data on key study parameters in the RWD source since observational studies are likely to fail in emulating trials when the data are not fit-for-purpose.[9–12]

Previous randomized controlled trial (RCT) emulation studies based on US claims data have been conducted for the use of antidiabetic drugs among individuals with type 2 diabetes (T2D).[6,13,14] This increases confidence that when used with appropriate design and analytic methods, these data can generate valid and robust causal conclusions. Similar studies in other disease areas have also been conducted using diverse RWD: France's National Health Data system and the United Kingdom's Clinical Practice Research Datalink.[8,15] However, there have not been comparable benchmarking exercises conducted with databases in Asia. With the growing amount of clinical evidence derived from large-scale Asian RWD and interest in its use in regulatory decision-making, such a benchmarking exercise could provide valuable insights regarding these data sources.

We aimed to evaluate the concordance of results from RWE studies conducted using two Asian RWD sources with the results of six cardiovascular outcome trials (CVOTs) of antidiabetic drugs in patients with T2D. Each RWE study was explicitly designed to emulate the design and analysis plan of the corresponding CVOTs. We also explored the potential for ethnic diversity to drive discordance in results by comparing the RWE study results with the results of Asian subgroups in the RCTs.

MATERIALS AND METHODS

Data sources

We utilized two health insurance claims databases: Korea's National Health Insurance Service-National Health Insurance Database (NHID) (2010 through 2022), and Taiwan's National Health Insurance Research Database (NHIRD) (2011 through 2020). Both regions operate with single-payer healthcare systems that cover the health insurance of their entire residents (> 50 million for Korea, > 23 million for Taiwan), with continuous enrollment from birth (or enrollment) until death, without interruptions. This structure ensures high data consistency and strong population representativeness. These data sources encompass comprehensive information on diagnoses, prescriptions, and procedures, in both inpatient and outpatient settings. Diagnosis records are coded according to the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) and Ninth Revision (ICD-9); ICD-10 for the entire study period in NHID and from 2016 in NHIRD, and ICD-9 for the period prior to 2016 in NHIRD.[16,17] Detailed explanation for each database is provided in **eAppendix 1** in **Supplement 1**. In this study, each site had converted the local database to a common data model and executed the same analytics to get the results. To ensure consistency when defining comorbidity covariates, we manually mapped corresponding diagnoses between ICD 9 and ICD 10 systems by reviewing clinical definitions and selecting the most appropriate matches.

Study design

We emulated six CVOTs in patients with T2D. These included four CVOTs of dipeptidyl peptidase 4 inhibitors (DPP4i) (CAROLINA (cardiovascular outcome study of linagliptin vs. glimepiride in patients with type 2 diabetes),[18] CARMELINA (cardiovascular and renal microvascular outcome study with linagliptin in patients with type 2 diabetes mellitus),[19] TECOS (sitagliptin cardiovascular outcomes study),[20] and SAVOR-TIMI 53 (does saxagliptin reduce the risk of cardiovascular events when used alone or added to other diabetes medications)),[21] as well as two CVOTs of sodium-glucose cotransporter 2 inhibitors (SGLT2i) (DECLARE-TIMI 58 (multicenter trial to evaluate the effect of dapagliflozin on the incidence of cardiovascular events),[22] and EMPA-REG OUTCOME (BI 10773 (Empagliflozin) cardiovascular outcome event trial in type 2 diabetes mellitus patients)).[23] The five ingredients (linagliptin, sitagliptin, saxagliptin, dapagliflozin, and empagliflozin), examined in the six CVOTs, were widely prescribed in both Korea and Taiwan during the study period, thereby ensuring a sufficient study population for robust analyses. Emulations for CANVAS (CANagliflozin cardioVascular Assessment Study)[24] and LEADER (Liraglutide effect and action in diabetes: evaluation of cardiovascular outcome results)[25] were originally planned but were not included in this study due to low usage or reimbursement considerations in Korea and Taiwan.

All emulations were conducted as an active comparator, new-user cohort study design, applying the same eligibility criteria with the corresponding CVOT.[26] The protocols for the RWE studies to emulate the trials were all pre-specified and registered on Open Science Framework (OSF) registries before the analysis (**eAppendix2** in **Supplement 1**). Protocols elaborated on operational definitions for eligibility criteria, exposure and comparators, outcome, and baseline characteristics for covariates using RWD. These elements were defined to be as analogous to trials as possible, while also taking into account the feasibility of implementing them from local data sources. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline (**Data S2**).

Study population

We included patients with T2D who initiated the drugs of interest in each CVOT. Assessment periods for eligibility criteria were defined specifically according to each criterion in each of the CVOTs as presented in **Tables S1–S6**, and in cases where no specific period was mentioned, we applied a 180-day period prior to the first prescription of study drugs (**Figures S1–S6**). We defined the index date as the date of the first prescription for the drug of interest during the study period. Patients were allowed to enter the cohort from the date of reimbursement of the drug to the end date of the available data (31 December 2022), except where the start date of the available data was after the reimbursement date.

Exposures and comparators

Among the six CVOTs, five evaluated the drug of interest (linagliptin, sitagliptin, saxagliptin, dapagliflozin, empagliflozin) compared to

placebo, and only CAROLINA evaluated linagliptin compared to glimepiride as an active comparator. Given that comparisons with non-users of antidiabetic drugs are likely to suffer from confounding by indication or time-lag bias in RWE studies,[27] we selected active comparators for each emulation as proxies for placebo. Glimepiride was selected as the active comparator in the emulation for CAROLINA and CARMELINA. Second generation sulfonylureas were selected as the active comparator in the emulations for the CVOT of other DPP4i (SAVOR-TIMI, TECOS). DPP4i were selected as the active comparator in the emulations for CVOTs of SGLT2i (DECLARE-TIMI 58, EMPA-REG OUTCOME). All study drugs and active comparators are in a similar line of therapy with "ntermediate" or "high" efficacy of glycemic management.[28]

Follow-up

Patients were followed up based on an on-treatment approach: they were followed from the index date until the earliest of outcome occurrence, switching to or adding study drug other than index drug, drug discontinuation, death, or the end of study period (31 December 2022). A 60-day grace period was applied to define drug discontinuation, regarding the uncertainty of adherence in clinical practice. As such, it was considered continuous treatment if a subsequent prescription was issued within 60 days after the end of the previous prescription's days' supply. Patients were censored 60 days after the end of the last prescription if the gap between prescriptions exceeded this period. Duration of follow-up for each of the treatment groups in each emulation is presented in **Table S7**.

Outcomes

We aimed to emulate the primary cardiovascular endpoint for the corresponding CVOT. For DECLARE-TIMI58, EMPA-REG OUTCOME, CARMELINA, CAROLINA, and SAVOR-TIMI53, the primary outcome was 3-point major adverse cardiovascular events (MACE) comprising nonfatal myocardial infarction, nonfatal stroke, and death from cardiovascular causes. For TECOS, in addition to 3P-MACE, unstable angina was included as the composite primary endpoint. The individual components of the composite primary endpoint were assessed as secondary outcomes. Cardiovascular death was defined as death with cardiovascular disease recorded as the primary cause of death. Several validation studies have been conducted within the NHID or NHIRD regarding these cardiovascular outcomes, as detailed in **Table S8**.

Statistical analysis

A range of covariates including age, sex, comorbidities, comedications, and proxies for health care utilization were measured over the baseline assessment period, defined as 180 days prior to and including the index date (**Table S9**). Propensity scores (PS) were estimated conditional on all the covariates using logistic regression models for each emulation. Covariate distributions were balanced between each treatment group via 1:1 PS matching based on the greedy nearest-neighbor method with a caliper of 0.1. We calculated the number of events and the incidence rate of outcome events per 100 person-years based on Poisson distribution in the PS-matched cohorts. PS-matched hazard ratios (HR) with 95% confidence intervals (CI) were estimated using Cox proportional hazard models. Additional adjustment was made in the Cox model only for covariates with an absolute standardized mean difference value > 0.1 after PS matching. We pooled the database-specific HR estimates from Korea and Taiwan using random-effects meta-analysis and assessed heterogeneity between databases using the I^2 statistic.[29] Statistical analyses were conducted using SAS software, version 9.4 (SAS Institute Inc., Cary, NC, USA) and R software, version 4.3.2 (R Project for Statistical Computing).

Assessment of agreements between RCT-RWE findings

We compared our individual and pooled estimates to the estimates from corresponding CVOT. To assess the potential for discordance in conclusions that might stem from ethnic diversity, we also compared the results from Asian subgroups in each of the CVOT vs. Asian RWD.

Three metrics of agreements, all of which were developed and employed by the Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology (RCT-DUPLICATE initiative),[30] were adopted to assess the differences between the estimates from RCT-RWE findings. First, estimate agreement (EA) can be met if the estimate from the RWE study falls within the 95% CI of the corresponding CVOT. Second, the absolute value of the standardized difference for effect estimates of each RCT-RWE study pair being smaller than 1.96 denoted standardized difference agreement (SD). We elaborated on the estimation of standardized differences in **eAppendix 3** in **Supplement 1**. Last, statistical significance agreement (SA) regards the extent to which RWE study can emulate the direction and statistical significance of the CVOT findings. In this study, we considered SA to be met if the upper boundary of the 95% CI of the HR from the emulation did not exceed the pre-specified non-inferiority margin of the corresponding CVOT. All CVOT being emulated targeted non-inferiority with a margin of 1.3, except for SAVOR-TIMI53 (superiority trial).

Sensitivity analyses

Control outcomes were pre-specified for each of CVOTs after considering the similarity of the confounding structure for the control outcomes with that of the outcome of interest. Control outcomes were used to evaluate the potential impact of unmeasured confounders.[31] For the CVOTs of DPP4i comparing gliptins with sulfonylureas, we selected "severe hypoglycemia" as a positive control outcome due to its well-established association with sulfonylurea,[32] expecting HR estimates < 1. As for the CVOTs of SGLT2i, we selected "diabetic ketoacidosis" and "hospitalization for heart failure" as positive control outcomes, expecting HR estimates > 1, and < 1, respectively.[33,34] End-stage renal disease was pre-specified as a negative control outcome in CARMELINA but did not yield enough events to estimate the HR. We also conducted an intention-to-treat analysis without censoring patients on drug discontinuation or switching and carried forward the initial treatment for 365 days to address potential informative censoring.

ETHICS STATEMENT

Ethical approval was obtained from the Institutional Review Board of Sungkyunkwan University (SKKU 2022-05-004), where the requirement of informed consent was waived as this study used anonymized administrative data and National Cheng Kung University Hospital (NCKU A-ER-111-197).

RESULTS

Study population

Baseline characteristics in our emulation and corresponding CVOTs were presented in **Table 1**. After applying the same eligibility and exclusion criteria, several important patient characteristics had notably different distributions when compared with the corresponding CVOT. The populations in the RWE studies for EMPA-REG OUTCOME (mean age 58.1 years for RWE vs. 63.1 years for RCT) and SAVOR-TIMI53 (60.3 years vs. 65.1 years) were younger and TECOS (68.6 years vs. 62.9 years) were older compared to the corresponding CVOT. All six emulations

Table 1 Baseline characteristics of RCTs and the corresponding emulation of RWE studies

Characteristics	DECLARE -TIMI58	EMPA-REG OUTCOME	CARMELINA	CAROLINA	TECOS	SAVOR -TIMI53
Age, years, mean ± SD						
Korea	64.2±8.2	58.5±12.4	67.8±11.7	63.0±11.2	68.6±9.4	60.4±10.8
Taiwan	63.1±8.1	55.9±12.5	65.7±13.2	60.7±11.1	68.2±10.6	59.8±11.1
RWE, Asia	64.0±8.2	58.1±12.4	66.0±13.0	62.8±11.2	68.6±9.5	60.3±10.9
RCT	64.0±6.8	63.1±8.7	65.9±9.1	64.1±9.5	62.9±9.1	65.1±8.6
Female, %						
Korea	37.5	41.8	33.1	47.1	40.8	42.3
Taiwan	34.9	38.3	44.9	45.1	37.2	44.5
RWE, Asia	37.1	41.3	43.0	46.9	40.4	42.6
RCT	37.4	28.6	37.1	40.0	29.3	33.1
Hypertension, %						
Korea	59.5	75.0	50.7	61.7	55.1	60.1
Taiwan	64.2	79.6	69.4	63.3	70.2	64.5
RWE, Asia	60.3	75.7	66.5	61.9	56.6	60.7
RCT	89.4	95.0[†]	91.0	90.1	≥78.8[†]	81.8
Heart failure, %						
Korea	0.7	3.4	10.4	3.0	5.9	2.6
Taiwan	0.6	6.4	10.2	3.8	10.8	3.4
RWE, Asia	0.7	3.9	10.2	3.1	6.3	2.7
RCT	10.1	10.1	26.8	4.5	18.0	12.8
Antidiabetic drugs, %						
Metformin						
Korea	92.9	94.7	64.8	77.0	73.3	76.1
Taiwan	97.6	96.6	78.0	78.0	80.5	63.0
RWE, Asia	93.6	95.0	76.0	77.1	74.0	74.2
RCT	82.0	74.0	54.6	83.2	81.6	69.6
Insulin						
Korea	18.3	17.3	31.1	11.6	18.4	6.5
Taiwan	16.8	17.1	19.3	15.5	34.7	8.6
RWE, Asia	18.0	17.2	21.1	12.1	20.0	6.8
RCT	40.9	48.2	58.0	-	23.2	7.2
Cardiovascular drugs, %						
ARB						
Korea	59.2	80.6	63.7	51.4	56.1	56.1
Taiwan	54.1	76.5	53.5	46.6	56.9	46.6
RWE, Asia	58.4	79.9	55.1	50.8	56.2	54.8
RCT	81.3	80.7	22.7	31.2	78.8	27.9
Beta blockers						
Korea	16.0	18.0	53.5	16.9	35.5	16.1
Taiwan	30.6	39.9	41.6	28.0	48.4	27.4
RWE, Asia	18.3	21.2	43.4	18.2	36.8	17.6
RCT	52.6	64.9	59.5	39.0	63.5	61.6
Statins						
Korea	67.8	72.9	81.6	54.8	73.7	66.7
Taiwan	70.1	72.4	54.0	55.5	67.3	55.6

(Continued)

Table 1 (Continued)

Characteristics	DECLARE -TIMI58	EMPA-REG OUTCOME	CARMELINA	CAROLINA	TECOS	SAVOR -TIMI53
RWE, Asia	68.2	72.8	58.2	54.9	73.1	65.2
RCT	75.0	77.0	71.9	64.6	79.9	78.3
Asian, %						
RCT	13.4	21.6	9.2	17.6	22.3	10.8

CARMELINA, cardiovascular and renal microvascular outcome study with linagliptin in patients with type 2 diabetes mellitus; CAROLINA, cardiovascular outcome study of linagliptin vs. glimepiride in patients with type 2 diabetes; CVD, cardiovascular disease; DECLARE-TIMI58, multicenter trial to evaluate the effect of dapagliflozin on the incidence of cardiovascular events; EMPA-REG OUTCOME, empagliflozin cardiovascular outcome event trial in type 2 diabetes mellitus patients; MI, myocardial infarction; RCT, randomized controlled trial; RWE, real-world evidence; SAVOR-TIMI53, does saxagliptin reduce the risk of cardiovascular events when used alone or added to other diabetes medication; TECOS, sitagliptin cardiovascular outcomes study.

indicated a lower proportion of hypertension, heart failure, and use of insulin and beta blockers for patients seen in clinical practice compared to CVOT participants. The proportions of Asian participants for the CVOTs ranged from 9.2% (CARMELINA) to 22.3% (TECOS).

Comparison between RCT-RWE findings

Incidence rates of 3-point MACE per 100 person-years were lower in all emulations (ranging from 0.6 to 3.6) compared to the corresponding CVOT (ranging from 2.1 to 5.8), except for TECOS. In the case of TECOS, the incidence rates observed in emulation (ranging from 5.5 to 9.2) were higher than those reported in the corresponding CVOT (approximately 3.6) (**Table 2**). Point estimates from all emulations (DECLARE-TIMI58: HR 0.64, 95% CI 0.57–0.72; EMPA-REG OUTCOME: 0.59, 0.52–0.67; CARMELINA: 0.98, 0.64–1.51; CAROLINA: 0.88, 0.82–0.94; TECOS: 0.91, 0.85–0.98; SAVOR-TIMI53: 0.64, 0.56–0.74) suggested stronger benefit than that of the corresponding CVOT, and all emulations except CARMELINA indicated superiority, whereas the corresponding CVOT only demonstrated non-inferiority (**Figure 1**). The two Asian RWDs were homogenous in four emulations ($I^2 = 0.0\%$), but CARMELINA ($I^2 = 80.5\%$) and SAVOR-TIMI53 ($I^2 = 11.1\%$) presented heterogeneity. Results for secondary outcomes of each of the six emulations were presented in **Tables S10–S15**.

We assessed the agreement between emulation results and their respective CVOTs by three agreement metrics (EA, SD, SA), as shown in **Figure 1**. All emulations met at least one metric in comparisons between results of RCT and Asian RWE findings (RCT vs. RWE, Asia in **Figure 1**). Notably, emulations for CAROLINA and TECOS satisfied all three metrics. EA metrics were met in four of six emulations (CARMELINA, CAROLINA, TECOS, and SAVOR-TIMI53), where the point estimates from the emulation fell within the 95% CI of the corresponding RCT. SD metrics were satisfied in CARMELINA (SD 0.17), CAROLINA (SD 1.31), and TECOS (SD 1.09), with the standardized difference between the estimates being < 1.96. SA metrics were met in all emulations except for CARMELINA, which presented a wide 95% CI with an upper boundary exceeding the non-inferiority margin of 1.3.

In comparisons for the Asians subgroup results from CVOTs and Asian RWE findings (RCT, Asia vs. RWE, Asia in **Figure 1**), EMPA-REG OUTCOME and CARMELINA presented more comparable results, satisfying all three metrics. CAROLINA and TECOS also met all three metrics as they were in comparison to the overall results of CVOT. SAVOR-TIMI53 met only one metric (SA) in comparison to Asian-specific RCT results, as was the case in comparison to the overall results of CVOT.

Sensitivity analyses

Results for severe hypoglycemia met expectations for CAROLINA, TECOS, and SAVOR-TIMI53 (**Figure 2**). Also, results for the hospitalization for heart failure met expectations for DECLARE-TIMI58 and EMPA-REG OUTCOME. However, they did not meet expectations for diabetic ketoacidosis, presenting null estimates. The results of the intention-to-treat analyses were consistent with the main findings across all six emulations (**Tables S16–S21**).

DISCUSSION

In this binational study using Korea and Taiwan RWD, we emulated six CVOTs of antidiabetic drugs in patients with T2D. Compared to results from CVOTs, each emulation met at least one of the pre-specified binary agreement metrics. Three out of six emulations met the EA and SD metrics, and five emulations met the SA metrics. Results from emulations were more concordant with Asian-specific results from CVOTs, with four emulations meeting all agreement metrics. The concordance in results demonstrates the ability of these databases to capture key study parameters relevant to T2D and supports the notion that when coupled with rigorous methods, these data can generate valid and robust causal conclusions about drug effects within the T2D therapeutic area.

A range of observational studies emulated landmark trials. The RCT-DUPLICATE initiative emulated 32 clinical trials using US claims data and publicly disclosed the methodologies and results of these emulations.[6,13] All 6 CVOTs included in our study were emulated in RCT-DUPLICATE, providing grounds for comparison. Interestingly, as in our study, emulation for SAVOR-TIMI53 met only one agreement metric and the point estimate was lower than that of trial. A study emulating the GRADE trial (A comparative effectiveness study of major glycemia-lowering medications for treatment of type 2 diabetes) presented comparable results to the trial, even though it was published before the trial results were made public.[14] For other therapeutic areas, a study that emulated

Table 2 Comparison of study sizes, incidence rates between RCTs and the corresponding emulation of RWE studies

			Exposure			Comparator		
Trial	Outcome	Study	Event	*N*	IR[a]	Event	*N*	IR[a]
DECLARE-TIMI58	3P-MACE	RWE, Korea	436	36,159	0.8	827	36,159	1.2
		RWE, Taiwan	38	6,995	0.6	70	6,995	1.0
		RCT, Asia	76	1,093	—	79	1,093	—
		RCT	756	8,582	2.3	803	8,578	2.4
EMPA-REG OUTCOME	3P-MACE	RWE, Korea	343	38,068	0.6	637	38,068	1.0
		RWE, Taiwan	42	6,684	0.7	64	6,684	1.0
		RCT, Asia	79	1,006	—	58	511	—
		RCT	490	4,687	3.7	282	2,333	4.4
CARMELINA	3P-MACE	RWE, Korea	70	1,921	2.6	77	1,921	3.6
		RWE, Taiwan	269	11,040	2.9	172	11,040	1.8
		RCT, Asia	40	307	—	40	333	—
		RCT	434	3,494	5.8	420	3,485	5.6
CAROLINA	3P-MACE	RWE, Korea	1,667	96,056	1.0	1,596	96,056	1.2
		RWE, Taiwan	141	12,190	1.3	132	12,190	1.3
		RCT, Asia	49	531	—	55	530	—
		RCT	356	3,023	2.1	362	3,010	2.1
TECOS	3P-MACE + angina	RWE, Korea	1,431	17,073	5.5	1,497	17,073	6.2
		RWE, Taiwan	130	1,865	7.6	127	1,865	9.2
		RCT, Asia	—	—	—	—	—	—
		RCT	613	8,280	3.6	609	8,212	3.6
SAVOR-TIMI53	3P-MACE	RWE, Korea	367	40,908	0.6	549	40,908	1.0
		RWE, Taiwan	64	6,646	1.0	80	6,646	1.4
		RCT, Asia	—	—	—	—	—	—
		RCT	839	7,257	4.1	851	7,266	4.2

CARMELINA, cardiovascular and renal microvascular outcome study with linagliptin in patients with type 2 diabetes mellitus; CAROLINA, cardiovascular outcome study of linagliptin vs. glimepiride in patients with type 2 diabetes; CVD, cardiovascular disease; DECLARE-TIMI58, multicenter trial to evaluate the effect of dapagliflozin on the incidence of cardiovascular events; EMPA-REG OUTCOME, empagliflozin cardiovascular outcome event trial in type 2 diabetes mellitus patients; MACE, major adverse cardiovascular events; MI, myocardial infarction; RCT, randomized controlled trial; RWE, real-world evidence; SAVOR-TIMI53, does saxagliptin reduce the risk of cardiovascular events when used alone or added to other diabetes medication; TECOS, sitagliptin cardiovascular outcomes study.
[a]Incidence rate (IR) per 100 person-years.

ORAL Surveillance trial (safety study of tofacitinib vs. tumor necrosis factor inhibitor in subjects with rheumatoid arthritis) using US claims data demonstrated concordant results with the RCT, when an "RCT duplicate cohort" was constructed by emulating the trial's eligibility criteria.[7]

Several challenges that can induce emulation differences were presented in our study. For instance, choosing the comparable patient group in RWD as a proxy for the placebo arm is challenging, given that these patients may differ in baseline characteristics, adherence, and other unmeasured confounders that could influence outcomes. Incomplete emulations for SAVOR-TIMI53 presented across different data sources (US and Asia) may result from an inadequate placebo proxy, as sulfonylureas are increasingly prescribed to older individuals and those with lower socioeconomic status,[35,36] which may leave residual confounding. In addition, the clinical practice populations for EMPA-REG OUTCOME and SAVOR-TIMI53 were, on average, 5 years younger and had a 10% higher proportion of women than the CVOT participant populations. Comparison of baseline characteristics generally suggested greater disease severity (e.g., higher prevalence of hypertension, heart failure, use of insulin, and beta blockers) in the CVOT populations compared to the emulation cohorts, as reflected by the lower incidence rates of 3P-MACE in the emulations, except for TECOS.

When emulating both the CARMELINA and CAROLINA trials, we compared linagliptin with glimepiride. However, their comparability with the corresponding trials appears to differ, likely due to two factors: the comparator drug and study population. CAROLINA directly used glimepiride for the control arm, allowing a straightforward emulation, while in CARMELINA, we used glimepiride as a proxy for placebo. Additionally, CARMELINA focused on patients at high CV and renal risk, whereas CAROLINA had a broader population, without restricting participants to those with underlying renal risks. The CARMELINA trial defined patients with renal risk using the urine albumin-creatinine ratio and albuminuria, but we emulated this using only diagnostic

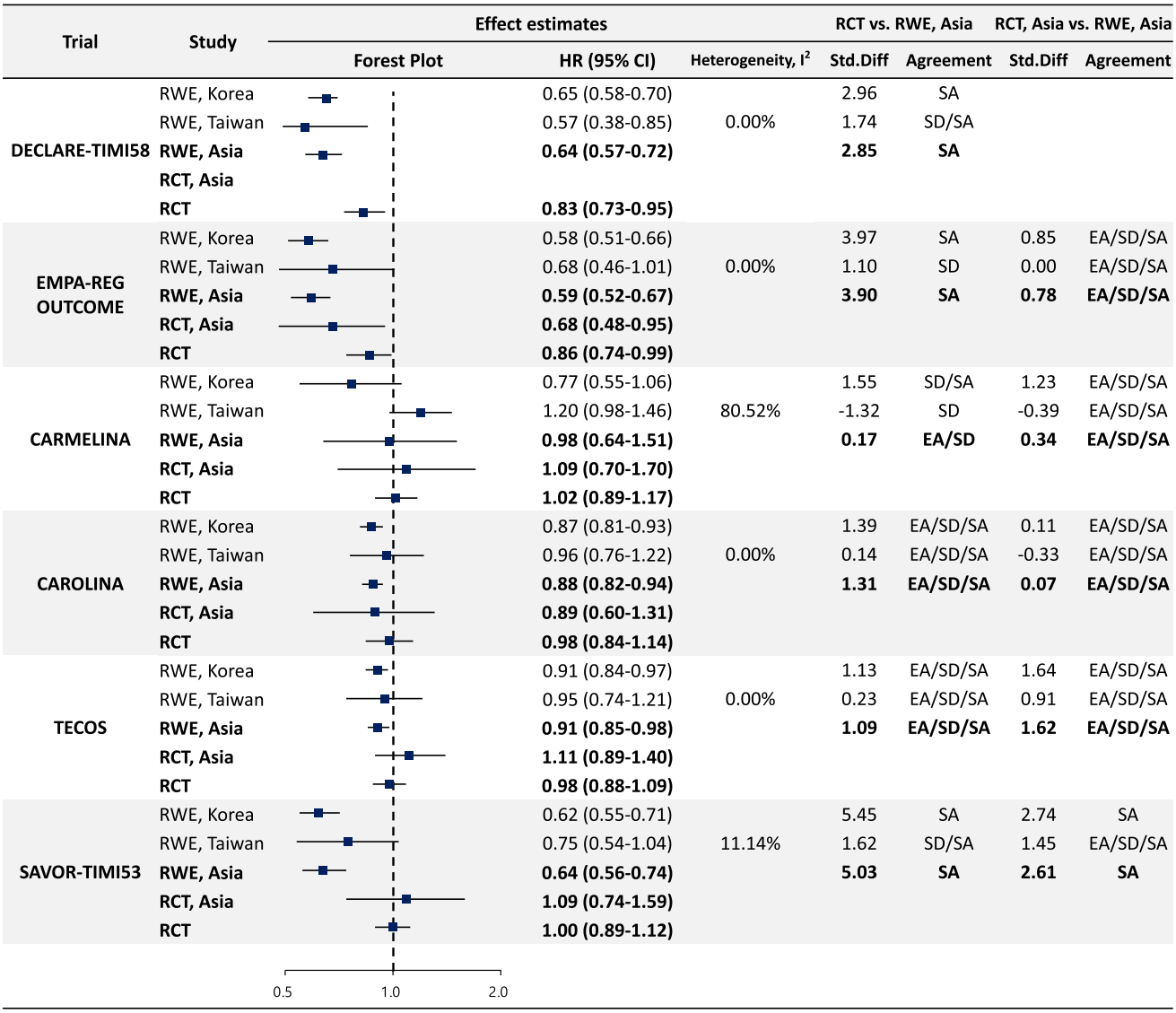

Figure 1 Effect estimates and evaluation of agreement between results from RCTs and the corresponding emulation of RWE studies. CARMELINA, cardiovascular and renal microvascular outcome study with linagliptin in patients with type 2 diabetes mellitus; CAROLINA, cardiovascular outcome study of linagliptin vs. glimepiride in patients with type 2 diabetes; CI, confidence interval; DECLARE-TIMI58, multicenter trial to evaluate the effect of dapagliflozin on the incidence of cardiovascular events; EA, estimate agreement; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) cardiovascular outcome event trial in type 2 diabetes mellitus patients; HR, hazard ratio; RCT, randomized controlled trial; RWE, real-world evidence; SA, statistical significance agreement; SAVOR-TIMI53, does saxagliptin reduce the risk of cardiovascular events when used alone or added to other diabetes medications; SD, standardized difference agreement; TECOS, sitagliptin cardiovascular outcomes study.

codes. This limitation likely contributed to our emulation of CAROLINA showing greater comparability than CARMELINA.

Our emulation results aligned more closely with the findings from the Asian subgroup of the corresponding CVOTs, suggesting that the ethnic differences observed in trials are supported by larger population-based observational studies. In the context of T2D, Asian patients tend to have a lower body mass index, a higher prevalence of beta-cell dysfunction,[37] earlier disease onset,[38] and distinct CV risk profiles compared to other ethnicities.[39] To generalize RCT findings across diverse populations, it is crucial to enhance the participation of underrepresented populations in these trials.[40,41] However, due to the limited diversity in many RCTs, Asian patients are often included only as small subgroups or conflicting meta-analyses.[42,43] This highlights the important role that well-designed observational studies using robust RWD play in addressing these gaps by generating evidence that is complementary to RCTs.

Strengths and limitations

To the best of our knowledge, this is the first study to emulate six CVOTs utilizing two large Asian data sources, demonstrating the ability of RWE studies using Asian RWD to generate similar

Trial	Outcome	Expected HR	Study	Exposure IR†	Comparator IR†	Observed HR (95% CI)	Met expectation
DECLARE-TIMI58	HF hospitalization	<1	RWE, Korea	1.2	3.5	0.98 (0.71-1.35)	Yes
			RWE, Taiwan	2.7	6.7	0.40 (0.23-0.70)	Yes
	Diabetic ketoacidosis	>1	RWE, Korea	1.3	1.1	0.98 (0.71-1.35)	No
			RWE, Taiwan	1.0	1.3	0.70 (0.25-1.98)	No
EMPA-REG OUTCOME	HF hospitalization	<1	RWE, Korea	2.1	4.5	0.46 (0.37-0.56)	Yes
			RWE, Taiwan	8.6	10.3	0.81 (0.57-1.15)	Yes
	Diabetic ketoacidosis	>1	RWE, Korea	1.2	1.4	0.79 (0.58-1.08)	No
			RWE, Taiwan	1.1	1.8	0.60 (0.24-1.53)	No
CAROLINA	Severe hypoglycemia	<1	RWE, Korea	3.2	8.2	0.41 (0.37-0.45)	Yes
			RWE, Taiwan	4.7	8.5	0.58 (0.41-0.82)	Yes
TECOS	Severe hypoglycemia	<1	RWE, Korea	5.2	7.2	0.72 (0.58-0.90)	Yes
			RWE, Taiwan	13.4	19.7	0.77 (0.45-1.33)	Yes
SAVOR-TIMI53	Severe Hypoglycemia	<1	RWE, Korea	1.5	5.8	0.25 (0.20-0.32)	Yes
			RWE, Taiwan	3.0	6.9	0.45 (0.26-0.78)	Yes

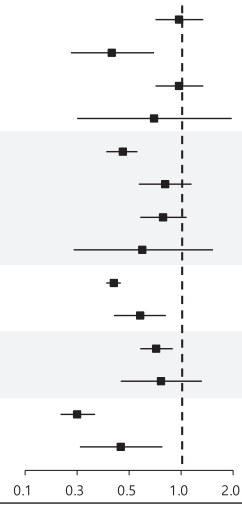

Figure 2 Results for control outcomes. CAROLINA, cardiovascular outcome study of linagliptin vs. glimepiride in patients with type 2 diabetes; CI, confidence interval; DECLARE-TIMI58, multicenter trial to evaluate the effect of dapagliflozin on the incidence of cardiovascular events; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) cardiovascular outcome event trial in type 2 diabetes mellitus patients; HF, heart failure; HR, hazard ratio; IR, incidence rate; SAVOR-TIMI53, does saxagliptin reduce the risk of cardiovascular events when used alone or added to other diabetes medications; TECOS, sitagliptin cardiovascular outcomes study. †Per 1,000 person-years.

conclusions as RCTs for a variety of glucose-lowering therapies. The two Asian RWD utilized in our study possess potential advantages for emulating trials. Since the single-payer health insurance system covers approximately 97% of the entire population, the target population of RWE studies conducted with this data source is highly representative of the population in real-world clinical practice, where the results of the RWE studies will be applied.[44,45] Moreover, the continuity of diagnostic and prescription records both within- and out-of-hospital ensures data completeness, thereby effectively capturing key study parameters such as exposures and outcomes.[46] Lastly, when implementing a common study protocol across different data sources, discrepancies may arise due to differences in the operational definitions of covariates, variations in statistical analysis methods, and the analytic infrastructure specific to each dataset. To minimize these differences, we not only developed a harmonized protocol but also executed unified SAS program analytics across both datasets.

There are some limitations to consider when interpreting our results. First, since the individuals in observational studies cannot be randomly assigned, comparability between treatment groups was achieved by PS matching. However, we were not able to measure potentially important risk factors due to limitations of the data source (e.g., missing for hemoglobin A1c levels, duration of diabetes), which leaves the potential for unmeasured confounding. Null estimates resulted from the positive control outcome of diabetic ketoacidosis for SGLT2i also suggest this potential. Caution is warranted when interpreting the risk estimates from our study. Second, the lack of information on the Asian population in published trials presents a limitation in interpreting our results. Lastly, the initially planned emulations of CVOTs for certain antidiabetic drugs, such as canagliflozin (CANVAS) and liraglutide (LEADER), were not conducted due to limited use or lack of reimbursement for these drugs, narrowing the scope of this study.

CONCLUSION

We observed comparable results between six database studies and the CVOTs they were designed to emulate. This comparability indicates that when coupled with rigorous design and analysis methods, database studies conducted with RWD from Korea and Taiwan can provide important complementary evidence to RCTs, especially for underrepresented Asian populations in RCTs, thereby supporting decision-making in both clinical and regulatory settings.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

FUNDING

This research was supported by a grant (RS-2024-00393167) from the Ministry of Food and Drug Safety, South Korea, in 2024–2028.

CONFLICTS OF INTEREST

J.-Y.S. received grants from the Ministry of Food and Drug Safety, the Ministry of Health and Welfare, the National Research Foundation of Korea, and pharmaceutical companies, including LG Chem, UCB, Pfizer, Celltrion, and SK Bioscience, outside the submitted work. All other authors declared no competing interests for this work. The funders had no role in the study design, collection, analysis, interpretation of data, writing of the report, and the decision to submit the article for publication.

AUTHOR CONTRIBUTIONS

H.Y.K. and S.P. wrote the manuscript. All authors designed the research. All authors performed the research. S.P., H.Y.K., A.C., H.E.J., B.M.-H.L., M.C.-Y.C., and P.C.M.A. analyzed the data. S.P. and J.-Y.S. contributed new reagents/analytical tools.

DATA AVAILABILITY STATEMENT

No additional data is available to the public.

1. Sherman, R.E. *et al*. Real-world evidence – what is it and what can it tell us? *N. Engl. J. Med.* **375**, 2293–2297 (2016).
2. Eichler, H.G. *et al*. Randomized controlled trials versus real world evidence: neither magic nor myth. *Clin. Pharmacol. Ther.* **109**, 1212–1218 (2021).
3. European Medicines Agency. Real-world evidence framework to support EU regulatory decision-making: 2nd report on the experience gained with regulator-led studies from February 2023 to February 2024. EMA, 180299 (2024).
4. US Food and Drug Administration. Framework for FDA's real-world evidence program <https://www.fda.gov/media/120060/download>. (2018) Accessed August 1, 2024.
5. Wang, S.V. & Schneeweiss, S. Assessing and interpreting real-world evidence studies: introductory points for new reviewers. *Clin. Pharmacol. Ther.* **111**, 145–149 (2022).
6. Franklin, J.M., Patorno, E., Desai, R.J. *et al*. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. *Circulation* **143**, 1002–1013 (2021).
7. Khosrow-Khavar, F., Kim, S.C., Lee, H., Lee, S.B. & Desai, R.J. Tofacitinib and risk of cardiovascular outcomes: results from the safety of TofAcitinib in routine care patients with rheumatoid arthritis (STAR-RA) study. *Ann. Rheum. Dis.* **81**, 798–804 (2022).
8. Wing, K. *et al*. Real world effects of COPD medications: a cohort study with validation against results from randomised controlled trials. *Eur. Respir. J.* **57**, 2001586 (2021).
9. Franklin, J.M. *et al*. When can nonrandomized studies support valid inference regarding effectiveness or safety of new medical treatments? *Clin. Pharmacol. Ther.* **111**, 108–115 (2022).
10. Franklin, J.M., Liaw, K.L., Iyasu, S., Critchlow, C.W. & Dreyer, N.A. Real-world evidence to support regulatory decision making: new or expanded medical product indications. *Pharmacoepidemiol. Drug Saf.* **30**, 685–693 (2021).
11. Dreyer, N.A. Advancing a framework for regulatory use of real-world evidence: when real is reliable. *Ther. Innov. Regul. Sci.* **52**, 362–368 (2018).
12. Simon, G.E. *et al*. When can we trust real-world data to evaluate new medical treatments? *Clin. Pharmacol. Ther.* **111**, 24–29 (2022).
13. Wang, S.V. *et al*. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* **329**, 1376–1385 (2023).
14. Deng, Y. *et al*. Emulating the GRADE trial using real world data: retrospective comparative effectiveness study. *BMJ* **379**, e070717 (2022).
15. Kirchgesner, J., Desai, R.J., Beaugerie, L., Kim, S.C. & Schneeweiss, S. Calibrating real-world evidence studies against randomized trials: treatment effectiveness of infliximab in Crohn's disease. *Clin. Pharmacol. Ther.* **111**, 179–186 (2022).
16. Hsieh, C.Y. *et al*. Taiwan's National Health Insurance Research Database: past and future. *Clin. Epidemiol.* **11**, 349–358 (2019).
17. Tsai, D.H. *et al*. Cross-regional data initiative for the assessment and development of treatment for neurological and mental disorders. *Clin. Epidemiol.* **15**, 1241–1252 (2023).
18. Rosenstock, J. *et al*. Effect of Linagliptin vs glimepiride on major adverse cardiovascular outcomes in patients with type 2 diabetes: the CAROLINA randomized clinical trial. *JAMA* **322**, 1155–1166 (2019).
19. Rosenstock, J. *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* **321**, 69–79 (2019).
20. Green, J.B. *et al*. Effect of Sitagliptin on cardiovascular outcomes in type 2 diabetes. *N. Engl. J. Med.* **373**, 232–242 (2015).
21. Scirica, B.M. *et al*. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N. Engl. J. Med.* **369**, 1317–1326 (2013).
22. Wiviott, S.D. *et al*. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N. Engl. J. Med.* **380**, 347–357 (2019).
23. Zinman, B. *et al*. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N. Engl. J. Med.* **373**, 2117–2128 (2015).
24. Neal, B. *et al*. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N. Engl. J. Med.* **377**, 644–657 (2017).
25. Marso, S.P. *et al*. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N. Engl. J. Med.* **375**, 311–322 (2016).
26. Lund, J.L., Richardson, D.B. & Stürmer, T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr. Epidemiol. Rep.* **2**, 221–228 (2015).
27. Suissa, S. & Azoulay, L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care* **35**, 2665–2673 (2012).
28. 9. Pharmacologic approaches to glycemic treatment: standards of Care in Diabetes-2024. *Diabetes Care* **47**(Suppl 1), S158–S178 (2024).
29. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
30. Franklin, J.M. *et al*. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin. Pharmacol. Ther.* **107**, 817–826 (2020).
31. Desai, R.J. *et al*. Process guide for inferential studies using healthcare data from routine clinical practice to evaluate causal effects of drugs (PRINCIPLED): considerations from the FDA sentinel innovation center. *BMJ* **384**, e076460 (2024).
32. Eriksson, J.W., Bodegard, J., Nathanson, D., Thuresson, M., Nyström, T. & Norhammar, A. Sulphonylurea compared to DPP-4 inhibitors in combination with metformin carries increased risk of severe hypoglycemia, cardiovascular events, and all-cause mortality. *Diabetes Res. Clin. Pract.* **117**, 39–47 (2016).
33. Liu, J. *et al*. Sodium-glucose co-transporter-2 inhibitors and the risk of diabetic ketoacidosis in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. *Diabetes Obes. Metab.* **22**, 1619–1627 (2020).
34. Kosiborod, M. *et al*. Lower risk of heart failure and death in patients initiated on sodium-glucose Cotransporter-2 inhibitors versus other glucose-lowering drugs: the CVD-REAL study (comparative effectiveness of cardiovascular outcomes in new users of sodium-glucose Cotransporter-2 inhibitors). *Circulation* **136**, 249–259 (2017).
35. Shin, H., Schneeweiss, S., Glynn, R.J. & Patorno, E. Trends in first-line glucose-lowering drug use in adults with type 2 diabetes in light of emerging evidence for SGLT-2i and GLP-1RA. *Diabetes Care* **44**, 1774–1782 (2021).
36. Lim, L.L. *et al*. Real-world usage of sulphonylureas in Asian patients with type 2 diabetes using the joint Asia diabetes evaluation (JADE) register. *Diabetes Obes. Metab.* **25**, 208–221 (2023).
37. Møller, J.B. *et al*. Body composition is the main determinant for the difference in type 2 diabetes pathophysiology between Japanese and Caucasians. *Diabetes Care* **37**, 796–804 (2014).
38. Lee, J.W., Brancati, F.L. & Yeh, H.C. Trends in the prevalence of type 2 diabetes in Asians versus whites: results from the United States National Health Interview Survey, 1997-2008. *Diabetes Care* **34**, 353–357 (2011).
39. Ueshima, H. *et al*. Cardiovascular disease and risk factors in Asia: a selected review. *CirculationCirculation* **118**, 2702–2709 (2008).
40. Schwartz, A.L., Alsan, M., Morris, A.A. & Halpern, S.D. Why diverse clinical trial participation matters. *N. Engl. J. Med.* **388**, 1252–1254 (2023).

41. US Food and Drug Administration. Diversity action plans to improve enrollment of participants from underrepresented populations in clinical studies guidance for industry <https://www.fda.gov/media/179593/download>. (2024) Accessed August 1, 2024.
42. Gan, S. *et al*. Efficacy of modern diabetes treatments DPP-4i, SGLT-2i, and GLP-1RA in white and Asian patients with diabetes: a systematic review and meta-analysis of randomized controlled trials. *Diabetes Care* **43**, 1948–1957 (2020).
43. Singh, A.K. & Singh, R. Cardiovascular outcomes with SGLT-2 inhibitors and GLP-1 receptor agonist in Asians with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Diabetes Metab. Syndr.* **14**, 715–722 (2020).
44. Cheol Seong, S. *et al*. Data resource profile: the National Health Information Database of the National Health Insurance Service in South Korea. *Int. J. Epidemiol.* **46**, 799–800 (2017).
45. Lin, L.Y., Warren-Gash, C., Smeeth, L. & Chen, P.C. Data resource profile: the National Health Insurance Research Database (NHIRD). *Epidemiol. Health* **40**, e2018062 (2018).
46. Schneeweiss, S. & Patorno, E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr. Rev.* **42**, 658–690 (2021).