Randomized Clinical Trial Replication Using Real-World Data for Comparative Cardiovascular and Safety Outcomes of Empagliflozin and Sitagliptin

Ha Young Jang

College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University

In-Wha Kim

College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University

Jung Mi Oh (✉ jmoh@snu.ac.kr)

College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University

Research Article

Keywords: Real-world evidence, Randomized controlled trials, Trial emulation, Diabetes, Sitagliptin, Empagliflozin

Posted Date: February 18th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1362085/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

If the same results as randomized clinical trial (RCT) can be created by conducting RWE studies, regulation decisions will be easier to assessing the effectiveness or safety of medications. In this study, we aimed to check to what extent findings from existing RCTs can be replicated in RWE study by analyzing in a reproducible process.

Methods

This study used a cohort study design to assess the effect of empagliflozin versus sitagliptin on effectiveness and safety outcomes. EMPA-REG OUTCOME trial (NCT01131676) was selected to target for replication. The inclusion/exclusion criteria and follow up method were applied as same with RCT. Major adverse cardiovascular events (MACEs) was used as a primary outcome and 15 other outcomes were also included for analysis.

Results

We followed-up (median, 2.7 years) 23,126 matched patients with type 2 diabetes mellitus (11,563 empagliflozin users and 11,563 sitagliptin users). Empagliflozin were related with a significantly decreased risk of MACEs [adjusted HR 0.87, 95% confidence interval (CI) 0.79–0.96]. The predefined estimate agreement, regulatory agreement, and standardized difference for RCT duplication were achieved. [EMPA-REG OUTCOME RCT: adjusted HR 0.86, 95% (CI) 0.74–0.99]. Of the 15 other emulation outcomes, 10 emulations were evaluated as good (all three agreements achieved), and two as moderate (two of the three agreements achieved).

Conclusions

Our study results suggest that RWE can replicate RCT results satisfactorily and have the potential for providing evidence for future regulatory decision-making. Further research is needed to determine whether RWE can be a reliable evidence in various clinical settings or specific patient groups.

Background

Randomized controlled trials (RCTs) are generally regarded as the gold standard for determining cause-effect relationships between medications and diseases; however, they have clear limitations: they are expensive, have a limited patient population, and are difficult to follow-up for a long time [(1, 2)]. Therefore, several attempts have been made to overcome these shortcomings. Real-world evidence (RWE) is clinical evidence concerning the potential benefits or risks of a medication derived from analysis of

real-world data (RWD). RWE has a relative advantage over RCTs because it enables a long-term follow-up study or research on rare populations. In the USA, the 21st Century Cures Act, passed in 2016, placed additional focus on the use of RWE to support regulatory decision making, including adding/modifying an indication, use in a new population, and adding comparative effectiveness or safety information [(3–6)]. With a rise in observational COVID-19 study dissemination, this trend is being accelerated [(7)]. To make regulatory decision-making through RWE possible, several RWE studies have been conducted and researchers have determined whether their results were the same as those of RCTs. However, direct comparison was difficult because the targeted population, outcomes, and treatment strategies were not exactly the same [(8)]. If the same results as those of RCTs can be created when the design and analysis method of the RWE study are implemented as closely as possible with the RCT, it will be easier to make regulatory decisions based on the results of future RWE studies. However, there have been few studies that have replicate the primary outcomes and other safety outcomes of the RCT.

The Food and Drug Administration requires a cardiovascular outcome trial (CVOT) to be conducted for a new diabetic medication to demonstrate its safety [(9)]. Various CVOTs investigating the use of diabetic medication have shown considerable clinical improvements, and regulatory submissions regarding safety information were made to several agencies [(10)]. Of those, the EMPA-REG OUTCOME study provided strong evidence that the sodium-glucose cotransporter 2 (SGLT2) inhibitor empagliflozin protects against major adverse cardiovascular events (MACEs) and other secondary outcomes [(11)]. This pivotal cardiovascular outcome trial is well-powered and used in a regulatory context.

In this study, we aimed to investigate the extent to which the findings from the EMPA-REG OUTCOME study can be replicated in an RWE study via a transparent and reproducible analysis process that would be acceptable for regulation. Furthermore, if there were any discrepancies between the RCT and RWE, we investigated the circumstances under which this inconsistency occurs.

Methods

Study Design and Sources

This was a cohort study to assess the effect of empagliflozin versus a placebo on cardiovascular and several safety outcomes of empagliflozin. The EMPA-REG OUTCOME trial (NCT01131676) [(11)] was selected to target replication [(12)]. The analyzed health insurance data was officially provided by the Korean Health Insurance Review & Assessment Service (HIRA) [(13)]. The insurance data included demographic, diagnosis, procedure, and prescription data of patients. The requirement for written informed consent from participants was waived because all participants were anonymized using a randomized identification number. This study was approved by the institutional review board of Seoul National University (IRB No. E2101/001-003). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline [(14)].

Study Patients

Patients who had been diagnosed with diabetes were included from 2011 to 2020. Patients were selected according to the same inclusion/exclusion criteria as a RCT (Additional file 1: Table S1). All patients (≥ 18 years) had established cardiovascular disease and received empagliflozin or sitagliptin for the first time. We selected an active comparator (sitagliptin) as a proxy for the placebo, because it is well known for observational studies, that a non-user comparator group can differ substantially from actively treated patients, unlike RCTs [(15)]. The index date was defined as the very first date each drug was prescribed.

Key Variables

MACEs outcome (all-cause death, myocardial infarction, and stroke) from the EMPA-REG OUTCOME trial was used as a primary outcome. An additional seven cardiovascular outcomes were analyzed: all-cause death, myocardial infarction (MI), hospitalization for unstable angina, coronary revascularization procedure, stroke, transient ischemic attack, and hospitalization for heart failure.

Eight safety outcomes were also analyzed: hypoglycemic events, urinary tract infections (UTIs), genital infections, volume depletion, acute kidney injury (AKI), diabetic ketoacidosis (DKA), thromboembolic events, and bone fracture. The operational definitions of outcomes were defined using the Korean Standard Classification of Diseases-9 codes or procedure codes and were directly matched to each Regulatory Activities Preferred Term (MedDRA PT) in the RCT (Additional file 1: Table S2). To minimize confounding variables (e.g., selection bias) as much as possible, 72 covariates were included viz. demographics, comorbidities, and disease/outcome specific variables. All covariates within the preceding 1 year of index date were evaluated.

Statistical Analysis

Statistical analyses were performed for the intention-to-treat population. Each time an outcome was analyzed, a new cohort was constructed after excluding patients with a history of the corresponding outcome. Patients were followed up until the earliest of events, the date of last follow up, the date of switching diabetic medication to the other comparison group, or the end of the study period. The maximum follow-up period was set at 48 months (same as in the RCT). Empagliflozin users were matched 1:1 to sitagliptin users and the distribution of the propensity score was inspected [(16)]. A standardized difference >0.1 was regarded as a sign of imbalance [(17)]. As same with RCT, the age and sex adjusted multivariate Cox proportional hazard regression was used to estimate the hazard ratio (HR) of empagliflozin for cardiovascular outcome, with a 95% confidence interval (CI). For a safety outcome model, logistic regression was used to the odds ratio (OR) of empagliflozin.

Sensitivity analyses were performed the same as with the RCTs in two ways. First, patients who received at least 1 dose of study drug were observed until ≤ 30 d after a patient's last intake of medication. Additionally, we followed up patients who received the study drug for ≥ 30 d (cumulative) including events that only occurred ≤ 30 d after a patient's last intake of medication ('as-treated' analysis). Analyses were performed with SAS Enterprise Guide version 7.1 (SAS Institute Inc., Cary, NC, USA).

RCT-RWE Agreement Assessment

We defined three metrics below to make a binary decision on whether an RCT was successfully replicated, considering statistical significance, directionality, and CIs associated with the corresponding RWE study [(12)]. First, we defined regulatory agreement as the ability of the RWE study to replicate the direction and statistical significance of the randomized trial finding. A secondary agreement metric was the estimate agreement, defined as an RWE estimate that lies within the RCT 95% CI. We also conducted hypothesis tests to evaluate whether there was a difference in findings by calculating the standardized difference (SD) between the RCT and RWE effect estimates. We considered a P value < 0.05 (where SD is greater than 1.96) statistically significant for the SD agreement. For comparison of results, HRs for cardiovascular outcomes and ORs for safety outcomes were calculated and compared (HRs were not provided for safety outcomes in an RCT). We defined the replication result as ‘good’ or ‘moderate’ if all three agreements or two of three agreements were achieved, respectively. If the replication result achieved $\leq$ one of the agreements, we defined the result as ‘fail’.

Results

Demographics

A total of 932,465 patients (age $\geq$ 18 years) diagnosed with diabetes who received empagliflozin or sitagliptin were identified. New empagliflozin or sitagliptin users (n = 384,579) were selected (Fig. 1). Among 98,733 patients who have high cardiovascular disease risks, an eligible study cohort with 48,545 patients remained after excluding patients who do not meet predefined inclusion criteria. Sitagliptin users were older and visited clinics more frequently (inpatient/outpatient) than empagliflozin users (Table 1). A later index date of empagliflozin users was observed compared to sitagliptin users. Compared to sitagliptin users, empagliflozin users were more often diagnosed with coronary artery disease (including coronary revascularization) and had fewer strokes.

After 11,563 empagliflozin users were matched to sitagliptin users, the above differences (age, number of clinic visits, index date, cardiovascular risk factors, comedications, and comorbidities) were reduced, and both groups were well balanced. Standardized differences were well below 0.1 for all 72 covariates. Median length of follow-up (2.7 years; median duration of anti-diabetic medications prescription during follow-up [1.7 (interquartile range 0.5–2.4) years]; and mean age of patients [55.6 years; men: 58.9% (n = 13,628)] were shown. In other nine study cohorts for evaluating safety outcomes, the two drug user groups were also well balanced after 1:1 matching. (Additional file 1: Tables S3-S11).

Table 1 Baseline characteristics

	Pre-Match			Post-Match		
Variables	Sitagliptin	Empagliflozin	STD	Sitagliptin	Empagliflozin	STD
	N= 36,861	N= 11,684		N= 11,563	N= 11,563	
Sex	20289 (55)	6913 (59.2)	-0.04	6799 (58.8)	6829 (59.1)	-0.004
Age	60.4 ± 11.4	55.4 ± 11	-0.4	55.5 ± 11.3	55.6 ± 10.9	0.008
Insurance type						
Normal	34434 (93.4)	11065 (94.7)	0.06	10927 (94.5)	10950 (94.7)	0.01
Medicaid	2234 (6.1)	583 (5)		595 (5.2)	577 (5)	
No charge	193 (0.5)	36 (0.3)		41 (0.4)	36 (0.3)	
Number of Inpatient visit	0.8 ± 2.2	0.5 ± 1.2	-0.2	0.5 ± 1.1	0.5 ± 1.2	0.01
Number of outpatient visit	28.1 ± 27.5	25.5 ± 23.9	-0.1	25.3 ± 23.1	25.6 ± 23.9	0.01
Time since type 2 diabetes mellitus						
≤1 years	18197 (49.4)	5045 (43.2)	0.1	5043 (43.6)	5008 (43.3)	0.007
>1 to 5 years	16927 (45.9)	5941 (50.9)		5827 (50.4)	5866 (50.7)	
>5 years	1737 (4.7)	698 (6)		693 (6)	689 (6)	
Index year						
2016	10568 (28.7)	2247 (19.2)	0.2	2281 (19.7)	2247 (19.4)	0.01
2017	18283 (49.6)	6336 (54.2)		6286 (54.4)	6270 (54.2)	
2018	8010 (21.7)	3101 (26.5)		2996 (25.9)	3046 (26.3)	
Charlson comorbidity index						
0	1521 (4.1)	473 (4.1)	0.03	469 (4.1)	468 (4.1)	0.004
1	3361 (9.1)	1144 (9.8)		1150 (10)	1138 (9.8)	
2	4937 (13.4)	1644 (14.1)		1634 (14.1)	1629 (14.1)	
3	27042 (73.4)	8423 (72.1)		8310 (71.9)	8328 (72)	

CV risk factor						
CAD	32597 (88.4)	10817 (92.6)	0.1	10657 (92.2)	10699 (92.5)	0.01
Multi vessel CAD	16230 (44)	6161 (52.7)	0.2	5944 (51.4)	6056 (52.4)	0.02
MI	1920 (5.2)	911 (7.8)	0.1	836 (7.2)	877 (7.6)	0.01
CABG	7665 (20.8)	3341 (28.6)	0.2	3195 (27.6)	3264 (28.2)	0.01
Stroke	5299 (14.4)	1062 (9.1)	-0.2	1030 (8.9)	1057 (9.1)	0.008
PAD	2376 (6.5)	666 (5.7)	-0.03	687 (5.9)	661 (5.7)	-0.010
DM circulation	4921 (13.4)	1802 (15.4)	0.06	1760 (15.2)	1775 (15.4)	0.004
DM foot	3 (0.0)	4 (0.0)	0.02	3 (0.0)	1 (0.0)	-0.01
DM nephropathy	2365 (6.4)	1009 (8.6)	0.08	965 (8.4)	985 (8.5)	0.006
DM neuropathy	5274 (14.3)	1591 (13.6)	-0.02	1620 (14)	1574 (13.6)	-0.01
DM other Complications	27326 (74.1)	8353 (71.5)	-0.06	8288 (71.7)	8287 (71.7)	0.000
Hyperglycemia	694 (1.9)	159 (1.4)	-0.04	160 (1.4)	154 (1.3)	-0.005
Comorbidities						
Hypertension	28872 (78.3)	9208 (78.8)	0.01	9095 (78.7)	9108 (78.8)	0.003
Edema	3490 (9.5)	1066 (9.1)	-0.01	1065 (9.2)	1056 (9.1)	-0.003
Kidney stone	585 (1.6)	168 (1.4)	-0.01	189 (1.6)	167 (1.4)	-0.02
Osteoarthritis	13169 (35.7)	3580 (30.6)	-0.1	3561 (30.8)	3568 (30.9)	0.001
Other arthritis	9104 (24.7)	2570 (22)	-0.06	2484 (21.5)	2554 (22.1)	0.02
PUD	9380 (25.5)	2828 (24.2)	-0.03	2733 (23.6)	2796 (24.2)	0.01
Pancreatitis	342 (0.9)	103 (0.9)	0.00	105 (0.9)	102 (0.9)	-0.003
UC	59 (0.2)	12 (0.1)	-0.02	10 (0.1)	12 (0.1)	0.006
Crohn	15 (0.0)	5 (0.0)	0.00	5 (0.0)	5 (0.0)	0.000
Asthma	5421 (14.7)	1626 (13.9)	-0.02	1596 (13.8)	1602 (13.9)	0.002
COPD	1349 (3.7)	296 (2.5)	-0.07	300 (2.6)	295 (2.6)	-0.003
Bladder stone	29 (0.1)	5 (0.0)	-0.01	5 (0.0)	5 (0.0)	0.000

Dementia	5993 (16.3)	1153 (9.9)	-0.2	1141 (9.9)	1150 (10)	0.003
Electrolyte Imbalance	2353 (6.4)	608 (5.2)	-0.05	585 (5.1)	600 (5.2)	0.006
Glaucoma /Cataract	10509 (28.5)	3176 (27.2)	-0.03	3127 (27)	3152 (27.3)	0.005
HONK	285 (0.8)	64 (0.6)	-0.03	65 (0.6)	63 (0.5)	-0.002
HTN nephropathy	166 (0.5)	54 (0.5)	0.00	40 (0.4)	51 (0.4)	0.02
Hyperthyroid disease	704 (1.9)	225 (1.9)	0.00	226 (2)	224 (1.9)	-0.001
Hypothyroid disease	1802 (4.9)	602 (5.2)	0.01	597 (5.2)	594 (5.1)	-0.001
Osteomyelitis	282 (0.8)	66 (0.6)	-0.02	56 (0.5)	66 (0.6)	0.01
Pneumonia	2872 (7.8)	770 (6.6)	-0.05	749 (6.5)	763 (6.6)	0.005
Skin infection	1438 (3.9)	459 (3.9)	0.00	459 (4)	455 (3.9)	-0.002
Glucose-lowering therapy						
Metformin	25836 (70.1)	8466 (72.5)	0.05	8422 (72.8)	8382 (72.5)	-0.008
Insulins	6312 (17.1)	2118 (18.1)	0.03	2098 (18.1)	2074 (17.9)	-0.005
SUs	16898 (45.8)	5499 (47.1)	0.02	5428 (46.9)	5441 (47.1)	0.002
Glitazones	3280 (8.9)	1328 (11.4)	0.08	1309 (11.3)	1301 (11.3)	-0.002
GLP-1 agonists	112 (0.3)	81 (0.7)	0.06	78 (0.7)	74 (0.6)	-0.004
AGIs	1532 (4.2)	364 (3.1)	-0.06	366 (3.2)	362 (3.1)	-0.002
Meglitinides	253 (0.7)	85 (0.7)	0.00	83 (0.7)	82 (0.7)	-0.001
Co-medications						
Anticoagulants	1650 (4.5)	564 (4.8)	0.02	523 (4.5)	550 (4.8)	0.01
Antiplatelets	24499 (66.5)	8232 (70.5)	0.09	8111 (70.2)	8141 (70.4)	0.006
Heparins	1287 (3.5)	354 (3)	-0.03	338 (2.9)	352 (3)	0.007
Thrombolytics	58 (0.2)	10 (0.1)	-0.02	7 (0.1)	10 (0.1)	0.01

Statins	25978 (70.5)	9459 (81)	0.3	9287 (80.3)	9343 (80.8)	0.01
Other lipid Lowerings	3903 (10.6)	1678 (14.4)	0.1	1627 (14.1)	1633 (14.1)	0.002
Nitrates	6264 (17)	2441 (20.9)	0.1	2378 (20.6)	2393 (20.7)	0.003
Digoxin	5390 (14.6)	2134 (18.3)	0.1	2060 (17.8)	2087 (18.1)	0.006
ACEIs	2127 (5.8)	963 (8.2)	0.1	929 (8)	927 (8)	-0.001
ARBs	21506 (58.3)	7292 (62.4)	0.08	7171 (62)	7198 (62.3)	0.005
Entresto	6 (0)	17 (0.2)	0.05	6 (0.1)	8 (0.1)	0.007
Other Anti HTNs	24131 (65.5)	8132 (69.6)	0.09	8017 (69.3)	8025 (69.4)	0.002
Loop diuretics	4310 (11.7)	1364 (11.7)	0.00	1292 (11.2)	1327 (11.5)	0.01
Other diuretics	10016 (27.2)	3223 (27.6)	0.01	3076 (26.6)	3165 (27.4)	0.02
Antianxieties	14982 (40.6)	4215 (36.1)	-0.09	4133 (35.7)	4183 (36.2)	0.009
Antipsychotics	1800 (4.9)	302 (2.6)	-0.1	299 (2.6)	301 (2.6)	0.001
Antidepressants	6667 (18.1)	1771 (15.2)	-0.08	1777 (15.4)	1759 (15.2)	-0.004
Dementia	5993 (16.3)	1153 (9.9)	-0.2	1141 (9.9)	1150 (10)	0.003
Antiparkinsons	1139 (3.1)	179 (1.5)	-0.1	164 (1.4)	179 (1.6)	0.01
Anticonvulsants	934 (2.5)	186 (1.6)	-0.07	200 (1.7)	186 (1.6)	-0.01
NSAIDs	28032 (76.1)	8810 (75.4)	-0.02	8757 (75.7)	8733 (75.5)	-0.005
Bisphos -phonates	1765 (4.8)	373 (3.2)	-0.08	379 (3.3)	371 (3.2)	-0.004
Opioids	16376 (44.4)	4778 (40.9)	-0.07	4720 (40.8)	4732 (40.9)	0.002

Values are represented as mean $\pm$ standard deviation or number (%); ACEIs, angiotensin-converting enzyme inhibitors; AGIs, α -glucosidase Inhibitors; ARBs, angiotensin II receptor blockers; CABG, coronary artery bypass graft; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; CV, cardiovascular; DM, diabetes mellitus; HONK, hyperglycaemic hyperosmolar nonketotic coma; HTN, hypertensive; MI, myocardial infarction; NSAIDs, non-steroidal anti-inflammatory drugs; PAD, peripheral

artery disease; PUD, peptic ulcer disease; STD, standardized difference; SUs, sulfonylureas; UC, ulcerative colitis;

(Location of Table 1)

Comparison of Baseline Characteristics Between RCT and RWE

A lower proportion of men and a lower mean age were observed in our RWE cohort than the corresponding RCT. (Table 2). Compared to the RCT, the RWE cohort was more often diagnosed with coronary artery disease (including coronary revascularization) and had fewer MIs, strokes, and peripheral artery disease. Rates of patients receiving glucose-lowering therapies were generally similar between the RCT and the RWE, except for the use of insulin. Patients in the RCT were more likely to have a longer time since diagnosis of type 2 diabetes than patients in RWE.

Table 2 Comparison of baseline characteristics between RCT and RWE

Values are represented as mean ± standard deviation or number (%); CV, cardiovascular; RCT, randomized clinical trial; RWE, real-world evidence

	EMPA-REG Outcome [®] (RCT)	EMPA-REG Duplicate (RWE)		
Characteristics	Placebo N= 2,333	Empagliflozin N= 4,687	Sitagliptin N= 11,563	Empagliflozin N= 11,563
Age	63.2 ± 8.8	63.1 ± 8.6	55.5 ± 11.3	55.6 ± 10.9
Male – no. (%)	1680 (72.0)	3336 (71.2)	6799 (58.8)	6829 (59.1)
CV risk factor				
Coronary artery disease	1763 (75.6)	3545 (75.6)	10657 (92.2)	10699 (92.5)
Multi-vessel coronary artery disease	1100 (47.1)	2179 (46.5)	5944 (51.4)	6056 (52.4)
History of myocardial infarction	1083 (46.4)	2190 (46.7)	836 (7.2)	877 (7.6)
Coronary artery bypass graft	563 (24.1)	1175 (25.1)	3195 (27.6)	3264 (28.2)
History of stroke	553 (23.7)	1084 (23.1)	1030 (8.9)	1057 (9.1)
Peripheral artery disease	479 (20.5)	982 (21.0)	687 (5.9)	661 (5.7)
Glucose-lowering therapy				
Metformin	1734 (74.3)	3459 (73.8)	8422 (72.8)	8382 (72.5)
Insulin	1135 (48.6)	2252 (48.0)	2098 (18.1)	2074 (17.9)
Sulfonylurea	992 (42.5)	2014 (43.0)	5428 (46.9)	5441 (47.1)
Thiazolidinedione	101 (4.3)	198 (4.2)	1309 (11.3)	1301 (11.3)
Glucagon-like peptide-1 agonist	70 (3.0)	126 (2.7)	78 (0.7)	74 (0.6)
Time since diagnosis of type 2 diabetes				
≤1 years	52 (2.2)	128 (2.7)	5043 (43.6)	5008 (43.3)
>1 to 5 years	371 (15.9)	712 (15.2)	5827 (50.4)	5866 (50.7)
>5 years	1910 (81.9)	3847 (82.1)	693 (6.0)	689 (6.0)
Anti-hypertensives	2221 (95.2)	4446 (94.9)	9625 (83.2)	9685 (83.8)
Diuretics	988 (42.3)	2047 (43.7)	3689 (31.9)	3698 (32.0)

Lipid-lowering	1864 (79.9)	3820 (81.5)	9616 (83.2)	9655 (83.5)
Anti-coagulants	2090 (89.6)	4162 (88.8)	8351 (72.2)	8422 (72.8)

Risk of Cardiovascular and Safety Outcomes

From the results of RWE, empagliflozin significantly reduced the risk of MACEs (HR 0.87, 95% CI 0.79–0.96), all-cause mortality (HR 0.78, 95% CI 0.67–0.91), and heart failure (HR 0.85, 95% CI 0.75–0.95) comparing to sitagliptin (Table 3). MI, stroke, hospitalization for unstable angina, coronary revascularization, and transient ischemic attack were not significantly associated with empagliflozin use.

For safety outcomes from RWE, empagliflozin was associated with lowered risk of hypoglycemia (HR 0.71, 95% CI 0.60–0.85) (Table 4), UTI (HR 0.89, 95% CI 0.83–0.95), AKI (HR 0.70, 95% CI 0.60–0.82), and volume depletion (HR 0.85, 95% CI 0.76–0.94) comparing to sitagliptin. Alternatively, the risk of genital infections significantly increased (HR 1.48, 95% CI 1.35–1.63) comparing to sitagliptin. No significant associations were identified in DKA, thromboembolic event, and fracture.

RCT-RWE Agreement

Regulatory agreement was found for 12 of 16 emulations (Table 3). The four emulations that showed disagreements were stroke, hypoglycemic adverse event, UTI, and volume depletion. For stroke, the HR estimate of RWE 0.89 was in the opposite direction to that of RCT. In RWE, empagliflozin showed significantly lowered risk, whereas the RCT reported a non-significant effect on hypoglycemic adverse event, UTI, and volume depletion.

An estimate agreement was achieved for 14 of the 16 emulations, with the exception for a hypoglycemic adverse event (OR: 0.70) and genital infections (OR: 1.48) where the emulation estimates were below the lower 95% CI bound from the RCT (OR: 1.00; 95% CI: 0.89–1.11 and OR: 3.74; 95% CI: 2.70–5.19 for hypoglycemic adverse event and genital infection, respectively) (Table 4). Statistically significant disagreements in SDs were shown (SD: -3.3 and -5.3 for hypoglycemic adverse event and genital infections, respectively).

Of the 16 emulation outcomes, 11 emulations were evaluated as good (all three agreements achieved), and three were moderate (two of three agreements achieved). Two emulation failed to replicate the RCT results (hypoglycemic adverse event and genital infections).

Table 3 RCT-RWE agreements for MACEs and each cardiovascular outcome component

Outcomes	EMPA-REG Outcome® (RCT)		EMPA-REG Duplicate (RWE)		STD	Agreement		
	Rate/1000	HR	Rate/1000	HR		RA	EA	SD
	Patient-yr	(95% CI)	Patient-yr	(95% CI)				
<i>MACEs</i>								
Sitagliptin	43.9	0.86	25.5	0.87	0.1	Y	Y	Y
Empagliflozin	37.4	(0.74-0.99)	22.5	(0.79-0.96)				
<i>All-cause death</i>								
Sitagliptin	28.6	0.68	12.0	0.78	1.0	Y	Y	Y
Empagliflozin	19.4	(0.57-0.85)	9.5	(0.67-0.91)				
<i>Myocardial infarction</i>								
Sitagliptin	19.3	0.87	8.7	0.91	0.3	Y	Y	Y
Empagliflozin	16.8	(0.70-1.09)	7.9	(0.76-1.08)				
<i>Stroke</i>								
Sitagliptin	10.5	1.18	9.1	0.89	-1.7	N	Y	Y
Empagliflozin	12.3	(0.89-1.56)	8.2	(0.75-1.05)				
<i>Hospitalization for unstable angina</i>								
Sitagliptin	10.0	0.99	50.5	0.94	-0.3	Y	Y	Y
Empagliflozin	10.0	(0.74-1.34)	48.1	(0.88-1.01)				
<i>Coronary revascularization</i>								
Sitagliptin	29.1	0.86	36.9	0.94	0.8	Y	Y	Y
Empagliflozin	25.1	(0.72-1.04)	35.2	(0.87-1.02)				
<i>Transient ischemic attack</i>								
Sitagliptin	3.5	0.85	9.2	0.88	0.1	Y	Y	Y
Empagliflozin	2.9	(0.51-1.42)	8.0	(0.74-1.04)				
<i>Hospitalization for heart failure</i>								
Sitagliptin	14.5	0.65	20.5	0.85	1.8	Y	Y	Y
Empagliflozin	9.4	(0.50-0.85)	17.4	(0.75-0.95)				

EA, estimate agreement; HR, hazard ratio; CI, confidence interval ; MACEs, major adverse cardiovascular events; RA, regulatory agreement, RCT, randomized clinical trial; RWE, real-world evidence; SD, standardized difference; STD, standardized difference; Y, yes; N, no

Table 4 RCT-RWE agreement for each safety outcomes

Outcomes	EMPA-REG Outcome® (RCT)		EMPA-REG Duplicate (RWE)		STD	Agreement		
	Rate (%)	OR (95% CI)	Rate (%)	OR (95% CI)		RA	EA	SD
<i>Hypoglycemic adverse event</i>								
Sitagliptin	27.9	1.00	2.6	0.70	-3.3	N	N	N
Empagliflozin	27.8	(0.89-1.11)	1.9	(0.59-0.84)				
<i>Urinary tract infection</i>								
Sitagliptin	18.1	0.99	23.3	0.87	-1.7	N	Y	Y
Empagliflozin	18.0	(0.87-1.13)	20.9	(0.81-0.94)				
<i>Genital infection</i>								
Sitagliptin	1.8	3.74	7.9	1.49	-5.3	Y	N	N
Empagliflozin	6.4	(2.70-5.19)	11.4	(1.35-1.65)				
<i>Acute kidney injury</i>								
Sitagliptin	6.6	0.78	3.3	0.70	-0.8	Y	Y	Y
Empagliflozin	5.2	(0.63-0.96)	2.3	(0.59-0.82)				
<i>Volume depletion</i>								
Sitagliptin	4.9	1.04	7.1	0.84	-1.7	N	Y	Y
Empagliflozin	5.1	(0.82-1.30)	6.1	(0.76-0.94)				
<i>Diabetic ketoacidosis</i>								
Sitagliptin	0.04	1.99	0.38	1.09	-0.5	Y	Y	Y
Empagliflozin	0.1	(0.2-17.8)	0.42	(0.72-1.64)				
<i>Thromboembolic event</i>								
Sitagliptin	0.9	0.75	4.3	0.92	0.7	Y	Y	Y
Empagliflozin	0.6	(0.42-1.31)	3.9	(0.80-1.05)				
<i>Fracture</i>								
Sitagliptin	3.9	0.98	13.8	0.94	-0.3	Y	Y	Y
Empagliflozin	3.8	(0.76-1.27)	13.1	(0.87-1.03)				

EA, estimate agreement; HR, hazard ratio; CI, confidence interval; MACEs, major adverse cardiovascular events; RA, regulatory agreement, RCT, randomized clinical trial; RWE, real-world evidence; SD, standardized difference; STD, standardized difference; Y, yes; N, no

Sensitivity Analyses

After follow-up of patients who received at least one dose of study drugs until ≤ 30 d after the last intake of medication, similar results (HR for MACEs: 0.88; 95% CI: 0.77–0.99) were obtained (Additional file 1: Table S12). Additional sensitivity analysis (including patients who received study drugs for ≥ 30 d including only events that occurred ≤ 30 d after a patient's last intake of medications) did not produce meaningful changes in the study findings (HR for MACEs: 0.87; 95% CI: 0.79–0.96) (Additional file 1: Table S13). All three agreements remained 'Y' for MACEs in both sensitivity analyses. In the same sensitivity analyses for eight safety outcomes, at least two of the three agreements were achieved in six safety outcomes (UTI, AKI, volume depletion, DKA, thromboembolic event, and fracture) (Additional file 1: Tables S14, S15). The hypoglycemic adverse event and genital infections still failed to show sufficient agreements, as in the main analysis.

Discussion

Our study analyzed patients with high cardiovascular disease risks that were prescribed empagliflozin or sitagliptin for emulation of a pre-existing RCT. The primary objective of the study was to evaluate the extent and reason for the differences between the original RCT and RWE results. To our knowledge, this is the first study to replicate the primary outcomes and other safety outcomes of the RCT. According to pre-specified agreement standards, 14 of the 16 RCT outcomes including primary outcomes were successfully reproduced (graded as 'good' or 'moderate'). Our study results suggested that RWE can replicate RCT results satisfactorily and have the potential for providing evidence for future regulatory decision-making.

To date, various RWE studies have reported on the efficacy and safety of SGLT-2 inhibitors including empagliflozin. There were discrepancies between findings, for example, the beneficial effect of SGLT-inhibitors on MACEs has been reported [(18–20)]. However, two other studies have reported non-significant results in MACEs [(21, 22)]. In other safety outcomes, Lega et al. reported a decreased risk of UTIs [(23)], while another study reported an association with an increased risk of UTI [(24)]. SGLT2 inhibitor use was associated with an elevated DKA risk [(25)]; however, this study was not in Korea [(26)]. We found both adverse [(27)] and beneficial [(28)] effects on fracture, although most results were non-significant. Most studies have reported decreased risks of SGLT2 inhibitors on AKI or impairment in renal function [(29–32)]. Those discrepancies may be due to differences in study samples, study designs, or statistical methods. To find a clear answer in a controlled clinical setting, reproducing an existing RCT design would be useful while securing the research quality of future RWE studies. In this way, we believe that researchers would predictably obtain reliable evidence.

Despite substantial effort, there were disagreements between the RCT and RWE in several outcomes. Stroke is a well-known disease that can be captured with a high accuracy because of its seriousness. The incidence rates were similar between RCT and RWE results. However, our study result suggested that empagliflozin was associated with a decreased risk of stroke (although not significant) unlike its non-significant increase in a RCT. Ethnic factors may have been involved because over 70% of patients were Caucasian, and only 20% were Asian in the RCT [(11)]. Asians are reported to have a lower risk of cardiovascular disease than other races [(33)]. As this study was conducted on Koreans, the proportion of patients with a history of severe diseases such as MI or stroke was small at baseline, and the age and severity of diabetes (time since onset of type 2 diabetes) were also lower than those of the RCT. In the subgroup analysis reported by a RCT, empagliflozin was reported to have a HR of 0.88 and 1.48 for Caucasians and Blacks for MACEs respectively and 0.68 for Asians [(11)]. Another study showed the protective effect of the SGLT-2 inhibitors against stroke in Koreans [(24)]; therefore, racial factors may have influenced our findings.

Another hypothesis includes a possibility of physicians' reluctance to prescribe empagliflozin because of its known side effects. The drug approval date of empagliflozin was May 2016 in Korea, and physicians may have paid attention to prescription in the early stages of approval during the index period (2016–2018) of this study. This trend is expected to be more prominent in outcomes such as stroke and genital infection in which the point estimate was reported as 1 or higher in RCTs. The HR point estimate of such an outcome in RWE is either reversed or much lower than the value reported in the RCT. Stroke and genital infection showed HRs of 1.18 and 3.59 in the RCT, and 0.89 and 1.49 in our RWE study, respectively. In fact, it has been reported that cardiologists may be reluctant to prescribe SGLT2 inhibitors owing to concerns of adverse effects [(34)]. Therefore, there is a possibility that undetected selection bias exists in our study.

In the hypoglycemic event, there was a > 10-fold difference between the incidence in a RCT and that in RWE. The hypoglycemic event was less likely to be captured in real world claim data, as shown in the event rates. It appears that physicians in Korea consider hypoglycemic events to be temporary and do not often insert a diagnostic code. Similarly, two observational studies in Korea showed low event rates of hypoglycemia (6.3% self-reported outcome) [(35)], and 2.4 per 100 person-year (insurance claim data)(24). The discrepancy in event rates could have led to the disagreement in treatment effect estimates. The event rate appears to be an important factor when conducting the RCT replication study.

The intention-to-treat approach was applied in our study, and the median duration of observation time was 2.7 and 3.1 years in RWE and RCT studies, respectively. Adherence to medications in the RWE is often poor compared with the RCT [(36)], and the median duration of treatment was 1.7 (RWE) and 2.6 (RCT) years in this study. In sensitivity analysis, as-treatment analyses were performed to test whether our main outcome was affected by adherence. Similar results were obtained, and shorter duration of use for empagliflozin provided a benefit on several outcomes.

There are several limitations in our study. We tried to emulate as much of a RCT as possible, including inclusion and exclusion criteria, exposures, and results; however, because of the limitations of the healthcare database, accurate emulation was not possible. Our study is a retrospective cohort design and not all information is included in the HIRA data (e.g., lab results for blood glucose test, urine culture test, or body weight). Therefore, although we adjusted for all possible confounders, there still may be residual confounding factors present. There were regulatory disagreements in UTIs and volume depletion outcomes, indicating potential for residual confounding factors related to these outcomes. The ultimate goal of our study was to utilize relevant RWE for regulatory decisions when no RCT evidence is available. The results of RCT and RWE are not always consistent. As mentioned above, event rates for testing specificity of outcome definition should be addressed. In addition, consideration of characteristics such as study participants, real-world clinical settings, and data availability might be important for enhancing the validity of study.

Conclusions

Our study results suggest that RWE can replicate RCT results satisfactorily and have the potential for providing evidence for future regulatory decision-making. Further research is needed to determine whether RWE findings can be reliable evidence in various clinical settings or specific patient groups.

Abbreviations

RCT

randomized controlled trial

RWE

real-world evidence

RWD

real-world data

CVOT

cardiovascular outcome trial

SGLT2

sodium-glucose cotransporter 2

MACEs

major adverse cardiovascular events

HIRA

health insurance review & assessment service

STROBE

strengthening the reporting of observational studies in epidemiology

MI

myocardial infarction

UTI

urinary tract infection
AKI
acute kidney injury
DKA
diabetic ketoacidosis
MedDRA PT
regulatory activities preferred term
HR
hazard ratio
CI
confidence interval
OR
odds ratio
SD
standardized difference.

Declarations

Ethics approval and consent to participate

The requirement for written informed consent from participants was waived because all participants were anonymized using a randomized identification number. This study was approved by the institutional review board of Seoul National University (IRB No. E2101/001-003).

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from Korean National Health Insurance Service but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Korean National Health Insurance Service.

Competing interests

Declarations of interest: none

Funding

This research was supported by a grant (20173MFDS171) from the Ministry of Food and Drug Safety in 2020.

Authors' contributions

HJ contributed to the conception and design of the study, data acquisition, analysis and interpretation of results, drafted, and revised the manuscript. JO and I-WK contributed to the conception and design of the study, analysis and interpretation of results, and revised the manuscript.

Acknowledgements

This study used HIRA research data (M20200708641) from the HIRA.

References

1. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. *Pharmacoeconomics*. 1999;15(5):423–34.
2. Garrison LP, Jr., Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health*. 2007;10(5):326–35.
3. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. 2018.
4. Food and Drug Administration. Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry 2018 [Available from: <https://www.fda.gov/media/99447/download>. Accessed 2021-09-24.]
5. Food and Drug Administration. Submitting Documents Utilizing Real-World Data and Real-World Evidence to FDA for Drugs and Biologics 2019 [Available from: <https://www.fda.gov/media/124795/download>. Accessed 2021-09-24.]
6. Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices 2019 [Available from: <https://www.fda.gov/media/99447/download>. Accessed 2021-09-24.]
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(10):1398–400.
8. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies: First Results From the RCT DUPLICATE Initiative. *Circulation*. 2021;143(10):1002–13.
9. Food and Drug Administration. Guidance for industry: diabetes mellitus—evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes 2008 [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diabetes-mellitus-evaluating-cardiovascular-risk-new-antidiabetic-therapies-treat-type-2-diabetes> Accessed 2021-09-24.]
10. Sharma A, Pagidipati NJ, Califf RM, McGuire DK, Green JB, Demets D, et al. Impact of Regulatory Guidance on Evaluating Cardiovascular Risk of New Glucose-Lowering Therapies to Treat Type 2 Diabetes Mellitus: Lessons Learned and Future Directions. *Circulation*. 2020;141(10):843–62.

11. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015;373(22):2117–28.
12. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clin Pharmacol Ther*. 2020;107(4):817–26.
13. Kim JA, Yoon S, Kim LY, Kim DS. Towards Actualizing the Value Potential of Korea Health Insurance Review and Assessment (HIRA) Data as a Resource for Health Research: Strengths, Limitations, Applications, and Strategies for Optimal Use of HIRA Data. *J Korean Med Sci*. 2017;32(5):718–28.
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495–9.
15. Food and Drug Administration. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Guidance for Industry and FDA Staff. 2013.
16. Parsons LS. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques. <http://www2.sas.com/proceedings/sugi26/p214-26.pdf> 2001 [Accessed 2021-09-24.]
17. NCSS-statistical-Software. Data Matching – Optimal and Greedy. http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Data_Matching-Optimal_and_Greedy.pdf [Accessed 2021-09-24.]
18. Persson F, Nystrom T, Jorgensen ME, Carstensen B, Gulseth HL, Thuresson M, et al. Dapagliflozin is associated with lower risk of cardiovascular events and all-cause mortality in people with type 2 diabetes (CVD-REAL Nordic) when compared with dipeptidyl peptidase-4 inhibitor therapy: A multinational observational study. *Diabetes Obes Metab*. 2018;20(2):344–51.
19. Filion KB, Lix LM, Yu OH, Dell'Aniello S, Douros A, Shah BR, et al. Sodium glucose cotransporter 2 inhibitors and risk of major adverse cardiovascular events: multi-database retrospective cohort study. *BMJ*. 2020;370:m3342.
20. Dave CV, Kim SC, Goldfine AB, Glynn RJ, Tong A, Paterno E. Risk of Cardiovascular Outcomes in Patients With Type 2 Diabetes After Addition of SGLT2 Inhibitors Versus Sulfonylureas to Baseline GLP-1RA Therapy. *Circulation*. 2021;143(8):770–9.
21. Norhammar A, Bodegard J, Nystrom T, Thuresson M, Nathanson D, Eriksson JW. Dapagliflozin and cardiovascular mortality and disease outcomes in a population with type 2 diabetes similar to that of the DECLARE-TIMI 58 trial: A nationwide observational study. *Diabetes Obes Metab*. 2019;21(5):1136–45.
22. Jeon JY, Ha KH, Kim DJ. Cardiovascular Safety of Sodium Glucose Cotransporter 2 Inhibitors as Add-on to Metformin Monotherapy in Patients with Type 2 Diabetes Mellitus. *Diabetes Metab J*. 2021;45(4):505–14.
23. Lega IC, Bronskill SE, Campitelli MA, Guan J, Stall NM, Lam K, et al. Sodium glucose cotransporter 2 inhibitors and risk of genital mycotic and urinary tract infection: A population-based study of older women and men with diabetes. *Diabetes Obes Metab*. 2019;21(11):2394–404.

24. Han SJ, Ha KH, Lee N, Kim DJ. Effectiveness and safety of sodium-glucose co-transporter-2 inhibitors compared with dipeptidyl peptidase-4 inhibitors in older adults with type 2 diabetes: A nationwide population-based study. *Diabetes Obes Metab*. 2021;23(3):682–91.
25. Wang L, Voss EA, Weaver J, Hester L, Yuan Z, DeFalco F, et al. Diabetic ketoacidosis in patients with type 2 diabetes treated with sodium glucose co-transporter 2 inhibitors versus other antihyperglycemic agents: An observational study of four US administrative claims databases. *Pharmacoepidemiol Drug Saf*. 2019;28(12):1620–8.
26. Kim YG, Jeon JY, Han SJ, Kim DJ, Lee KW, Kim HJ. Sodium-glucose co-transporter-2 inhibitors and the risk of ketoacidosis in patients with type 2 diabetes mellitus: A nationwide population-based cohort study. *Diabetes Obes Metab*. 2018;20(8):1852–8.
27. Ueda P, Svanstrom H, Melbye M, Eliasson B, Svensson AM, Franzen S, et al. Sodium glucose cotransporter 2 inhibitors and risk of serious adverse events: nationwide register based cohort study. *BMJ*. 2018;363:k4365.
28. Toulis KA, Bilezikian JP, Thomas GN, Hanif W, Kotsa K, Thayakaran R, et al. Initiation of dapagliflozin and treatment-emergent fractures. *Diabetes Obes Metab*. 2018;20(4):1070–4.
29. Nadkarni GN, Ferrandino R, Chang A, Surapaneni A, Chauhan K, Poojary P, et al. Acute Kidney Injury in Patients on SGLT2 Inhibitors: A Propensity-Matched Analysis. *Diabetes Care*. 2017;40(11):1479–85.
30. Cahn A, Melzer-Cohen C, Pollack R, Chodick G, Shalev V. Acute renal outcomes with sodium-glucose co-transporter-2 inhibitors: Real-world data analysis. *Diabetes Obes Metab*. 2019;21(2):340–8.
31. Heerspink HJL, Karasik A, Thuresson M, Melzer-Cohen C, Chodick G, Khunti K, et al. Kidney outcomes associated with use of SGLT2 inhibitors in real-world clinical practice (CVD-REAL 3): a multinational observational cohort study. *Lancet Diabetes Endocrinol*. 2020;8(1):27–35.
32. Koh ES, Han K, Nam YS, Wittbrodt ET, Fenici P, Kosiborod MN, et al. Renal outcomes and all-cause death associated with sodium-glucose co-transporter-2 inhibitors versus other glucose-lowering drugs (CVD-REAL 3 Korea). *Diabetes Obes Metab*. 2021;23(2):455–66.
33. Jung KJ, Jang Y, Oh DJ, Oh BH, Lee SH, Park SW, et al. The ACC/AHA 2013 pooled cohort equations compared to a Korean Risk Prediction Model for atherosclerotic cardiovascular disease. *Atherosclerosis*. 2015;242(1):367–75.
34. Vardeny O, Vaduganathan M. Practical Guide to Prescribing Sodium-Glucose Cotransporter 2 Inhibitors for Cardiologists. *JACC Heart Fail*. 2019;7(2):169–72.
35. Hong AR, Koo BK, Kim SW, Yi KH, Moon MK. Efficacy and Safety of Sodium-Glucose Cotransporter-2 Inhibitors in Korean Patients with Type 2 Diabetes Mellitus in Real-World Clinical Practice. *Diabetes Metab J*. 2019;43(5):590–606.
36. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*. 2013;347:f6409.

Figures

Figure 1

Study flow chart

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)