Heterogeneous Treatment Effects on Cardiovascular Diseases With Dipeptidyl Peptidase-4 Inhibitors Versus Sulfonylureas in Type 2 Diabetes Patients

Chen-Yi Yang[1], Wei-Ann Lin[2,†], Pei-Fang Su[2], Lun-Jie Li[1], Chun-Ting Yang[1], Huang-Tz Ou[1,3,4,*] and Shihchen Kuo[5,6]

This study explored heterogeneous treatment effects (HTEs) of the real-world use of dipeptidyl peptidase-4 inhibitors (DPP-4is) vs. sulfonylureas (SUs) on cardiovascular diseases (CVDs) and mortality in patients with type 2 diabetes. Utilizing Taiwan's National Health Insurance Research Database, 19,853 propensity score-matched pairs of DPP-4i and SU stable users were identified. Classification and regression tree analyses and Cox models were applied to explore HTEs, according to various patient characteristics, on the composite CVDs, three-point major adverse cardiovascular event (MACE), and all-cause mortality. The absolute risk difference (ARD), hazard ratio (HR), and 95% confidence interval (CI) were estimated for comparing treatment effects. CVD history, ischemic stroke, or transient ischemic attack (IS/TIA) history, and age at treatment initiation were significant treatment effect modifiers. Patients with prior IS/TIA but without any other prior CVDs benefited most in reduced risks of composite CVDs from using DPP-4i vs. SU (ARD −4.31%, 95% CI −7.48% to −1.14%, HR 0.81, 95% CI 0.69 ~ 0.95), followed by those without prior IS/TIA and CVDs and initiated with DPP-4i at age < 69.3 years (ARD −0.90%, 95% CI −1.47% to −0.32%, HR 0.86, 95% CI 0.77 ~ 0.97). Patients with prior IS/TIA benefited most in reduced risks of three-point MACE from using DPP-4i vs. SU (ARD −4.22%, 95% CV −6.66% to −1.78%, HR 0.80, 95% CI 0.69 ~ 0.93), followed by those without prior IS/TIA and initiated with DPP-4i at age < 69.3 years (ARD −0.68%, 95% CI −1.08% to −0.29%, HR 0.81, 95% CI 0.70 ~ 0.93). Consideration of CVD and IS/TIA histories and age could facilitate individualized diabetes management of using DPP-4i vs. SU. Future studies are warranted given the hypothesis-generating nature in this exploratory research.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ A large, randomized, active comparator trial revealed that the use of dipeptidyl peptidase-4 inhibitor (DPP-4i) in patients with type 2 diabetes was noninferior to sulfonylurea (SU) for the composite cardiovascular disease (CVD) outcome, whereas several real-world comparative effectiveness studies showed the significantly reduced CVD risks from DPP-4i vs. SU.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ Potential heterogeneous treatment effects (HTEs) suggested by inconsistent results of CVD risks with DPP-4i vs. SU were explored using machine learning-based analyses.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ CVD history, ischemic stroke or transient ischemic attack (IS/TIA) history, and age were important factors contributing to heterogeneous CVD effects of DPP-4i vs. SU. Patients with IS/TIA but without other CVD histories had up to a 20% CVD risk reduction by using DPP-4i, and those without IS/TIA and CVD histories and initiated with DPP-4i at age ≤ 69 years had up to a 19% CVD risk reduction.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ The findings of HTEs for DPP-4i vs. SU can facilitate individualized diabetes management.

[1]Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [2]Department of Statistics, National Cheng Kung University, Tainan, Taiwan; [3]Department of Pharmacy, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [4]Department of Pharmacy, National Cheng Kung University Hospital, Tainan, Taiwan; [5]Division of Metabolism, Endocrinology, and Diabetes, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA; [6]Michigan Center for Diabetes Translational Research, University of Michigan, Ann Arbor, Michigan, USA. *Correspondence: Huang-Tz Ou (huangtz@mail.ncku.edu.tw)

[†]Equal to the first authorship.

Received June 8, 2020; accepted September 9, 2020. doi:10.1002/cpt.2058

Dipeptidyl peptidase-4 inhibitors (DPP-4is) are commonly used as a second-line or third-line glucose-lowering agent (GLA).[1] They have shown moderate glucose-lowering efficacy with a low risk of hypoglycemia, a neutral effect on body weight, and lower drug acquisition costs compared with other newer drugs, such as glucagon-like peptide-1 (GLP-1) receptor agonists and SGLT2 inhibitors.[2] All of four large cardiovascular outcome trials for DPP-4i used a placebo-controlled design and did not show cardiovascular benefits relative to placebo.[3–6] The CAROLINA[7] was the only large, randomized, active comparator trial of DPP-4i (linagliptin) vs. sulfonylureas (SU; glimepiride), another commonly used GLA, and it revealed that the use of DPP-4i relative to SU did not yield beneficial cardiovascular effects either. Nevertheless, our literature review (**Table S1**) has found that numerous cohort studies[8–17] for assessing long-term comparative risks of cardiovascular diseases (CVDs) associated with the real-world use of DPP-4i vs. SU showed the significantly reduced risks of CVDs from DPP-4i vs. SU.

These inconsistent results of CVD outcomes associated with the use of DPP-4i between clinical trials and real-world studies have implied the importance of evaluating potential heterogeneous treatment effects (HTEs) of DPP-4i, which could be hidden in a diverse real-world patient population with type 2 diabetes (T2D). That is, treatment outcomes may differ across patient subgroups by demographic and clinical characteristics. The results of investigating HTEs would be critical in clinical care for determining personalized treatments and in healthcare policy decisions for formulating effective healthcare reimbursement policies, which can be used to prioritize treatment strategies to gain benefit and avoid harm for patients.

Against this background, this study explored the complexity of multiple clinical factors that could explain heterogeneous cardiovascular effects of DPP-4i vs. SU in real-world patients with T2D by using machine-learning approaches that have been increasingly applied to analyze HTEs of antidiabetic treatments.[18,19]

METHODS

The Institutional Review Board of National Cheng Kung University Hospital approved the study before commencement (B-EX-103-015).

Data source

This study utilized data from the National Health Insurance Research Database (NHIRD) recorded from 1996–2013, which comprised de-identified, individual-level, longitudinal claims data of medical and pharmacy records from Taiwan's National Health Insurance (NHI) program.[20] The NHI program is a mandatory-enrollment single-payment system that covers over 99% of the population of 23 million people in Taiwan.[21]

Cohort identification

The prevalence of T2D in people aged 20–79 years in Taiwan during 2000–2013 was from 4.3–9.8%.[22,23] In this study, patients aged 18 years or above and newly diagnosed with T2D were identified from the NHIRD 1999–2013 and those receiving DPP-4i or SU during 2011–2012 were further identified (**Figure S1**). Patients using the combination of DPP-4i and SU were excluded because the main interest of study was to assess the individual effect and HTEs of DPP-4i vs. SU. The period of 2011–2012 was chosen because the reimbursement of DPP-4i by the NHI program started since 2009 and the utilization of DPP-4i increased steadily after 2010 in Taiwan.[1] During 2011–2013, SU and DPP-4i were among the most frequently used second-line GLAs for the patients with T2D failing to metformin monotherapy.[1] We then identified stable users of study drugs who newly initiated DPP-4i (n = 66,400) or SU (n = 21,600) in 2011–2012 without any prescriptions of these drugs in the preceding year. To avoid potential confounding from short-term or accidental use of study drugs, we selected stable users of DPP-4i or SU; stable users were patients with at least 3 consecutive refills of DPP-4i or SU with any gaps between 2 consecutive refills of < 30 days.[11,24] The index date for each study subject was the first date of initiating DPP-4i or SU during 2011–2012.

The propensity score (PS) matching procedure of the 8-digit greedy, nearest-neighbor 1:1 matching was applied to minimize the potential selection bias and enhance the comparability of 2 treatment groups. The critical baseline patient characteristics shown in **Table 1** were considered in the PS matching procedure, and the standardized mean difference statistics was used to indicate a satisfactory balance in patient characteristics between treatment groups if an absolute standardized mean difference value was < 0.1.

Operational definitions of medications and outcomes

In the NHIRD, the exposure of GLAs and CVD-related medications (**Table 1**) was measured according to the World Health Organization Anatomical Therapeutic Chemical Classification system. Study outcomes included: (i) the composite CVD event consisting of fatal or nonfatal myocardial infarction, ischemic heart disease, heart failure, cerebrovascular disease, cardiogenic shock, sudden cardiac arrest, arteriosclerotic cardiovascular disease, or arrhythmia, (ii) three-point major adverse cardiovascular event (MACE), including nonfatal myocardial infarction or stroke, or fatal CVDs, and (iii) all-cause mortality. Cardiovascular events were confirmed using International Classification of Disease, 9th revision-Clinical Modification codes from inpatient or emergency department records of the NHIRD, and the mortality status was ascertained from inpatient department records (**Table S2**). The coding accuracy for study outcomes in the NHIRD has been documented elsewhere.[11,25–28]

Statistical analyses

Our main analyses were conducted for patients without prior severe hypoglycemia (primary cohort) under an intention-to-treat (ITT) scenario. The ITT scenario analyses considered each subject to be observed from the index date until loss to follow-up from the NHI program, occurrence of a study outcome, or the end date of the database (December 31, 2013), whichever came first. Classification and regression tree (CART) analysis was applied to classify heterogeneous patient subgroups stratified by underlying demographic and clinical characteristics who had substantially different treatment effects. The CART is a commonly used and powerful machine-learning method in which the data are split into several subgroups, namely a node or leaf, by input predictors (e.g., demographic or clinical characteristics) indicating the heterogeneous features between the subgroups. The CART utilizes the common impurity approaches (e.g., Gini index for a binary outcome variable or minimizing the sum squared error for a continuous outcome variable) to assess the appropriateness of the classification for subgroups with the most homogeneous value of a given outcome.[29,30] The CART is a predictive algorithm for classifying patient subgroups by partitioning the data and fitting a prediction model within each subgroup.[29,31] The CART is a well-developed technique for analyzing the right-censored survival data, especially incorporating time-dependent covariates into tree structure analyses. All patient characteristics (**Table 1**) were used in the CART analysis as the potential estimators/factors that partitioned the data by differentiating the hazards of study outcomes associated with the treatment status (DPP-4i vs. SU). The final sets of subgroups (as nodes at the bottom of

Table 1 Baseline patient characteristics of DPP-4i and SU users after propensity score matching

Characteristics[a,b]	Primary cohort (without adjustment for prior severe hypoglycemia)[c]		Secondary cohort (with adjustment for prior severe hypoglycemia)[c]	
	DPP-4i (*n* = 19,853)	1:1 matched SU (*n* = 19,853)	DPP-4i (*n* = 19,856)	1:1 matched SU (*n* = 19,856)
Age at index date, year, mean ± SD	60.65 ± 12.39	60.38 ± 12.78	60.73 ± 12.44	60.35 ± 12.77
Male, %	53.30	54.34	54.04	54.38
Diabetes duration,[d] year, mean ± SD	5.81 ± 3.9	5.69 ± 3.74	5.83 ± 3.91	5.69 ± 3.73
Glucose-lowering agent in the year before the index date, MPR, 0-1, mean ± SD				
Metformin	0.33 ± 0.41	0.31 ± 0.40	0.33 ± 0.41	0.31 ± 0.40
Meglitinide	0.10 ± 0.28	0.11 ± 0.28	0.10 ± 0.28	0.11 ± 0.28
Thiazolidinedione	0.09 ± 0.25	0.10 ± 0.27	0.09 ± 0.25	0.10 ± 0.27
Acarbose	0.10 ± 0.27	0.11 ± 0.28	0.10 ± 0.27	0.11 ± 0.28
Insulin	0.04 ± 0.19	0.05 ± 0.20	0.05 ± 0.19	0.05 ± 0.20
Diabetes-related complication, %				
Retinopathy	10.05	10.15	10.20	10.15
Nephropathy	16.63	16.43	16.83	16.55
Neuropathy	11.80	11.47	12.02	11.51
Peripheral vascular diseases	5.19	5.21	5.27	5.27
IS or TIA[e]	10.30	9.33	9.90	9.30
Other CVDs[e]	6.81	6.47	6.85	6.39
CVDs[e]	24.45	23.34	24.66	23.48
Metabolic complications	1.46	1.54	1.63	1.53
Hypertension	66.28	64.80	66.34	64.77
Hyperlipidemia	57.16	55.26	56.77	55.18
Severe hypoglycemia[f]	N/A	N/A	1.71	1.59
CIC category,(%				
Cancer	6.97	6.88	7.16	6.89
Gastrointestinal disease	27.40	26.60	26.80	26.75
Musculoskeletal disease	38.95	38.43	39.13	38.63
Pulmonary disease	10.39	10.00	10.11	9.84
Substance abuse complexity	1.55	1.69	1.68	1.67
Mental illness	10.76	10.52	11.03	10.53
CVD-related medication, %				
Lipid modifying agents	56.82	54.30	56.97	54.34
α-blockers	5.17	4.92	5.07	4.89
β-blockers	34.52	33.81	34.59	33.86
Agents acting on the renin-angiotensin system	56.85	54.74	56.82	54.83
Diuretics	19.68	19.70	20.11	19.79
CCBs	40.87	40.70	41.11	40.69
Anti-arrhythmics	2.45	2.20	2.34	2.23
Cardiac glycosides	2.52	2.36	2.43	2.38
Vasodilators used in cardiac diseases	13.65	12.68	13.60	12.86
Antiplatelets	38.56	36.89	37.96	37.04
Anticoagulants	1.62	1.47	1.59	1.48

CCBs, calcium channel blockers; CIC, chronic illness with complexity; CVD, cardiovascular disease; DPP-4i, dipeptidyl peptidase-4 inhibitor; IS, ischemic stroke; MPR, medication possession ratio; N/A, not available; SU, sulfonylurea; TIA, transient ischemic attack.
The history of severe hypoglycemia was not measured in the primary cohort.
[a]All patient characteristics after propensity score matching were comparable between the two treatment groups, as supported by the value of the absolute standardized mean difference (SMD) of < 0.1. [b]All patient characteristics were measured in the year prior to the index date (i.e., the first date of initiation of DPP-4i or SU during 2011–2012), except age and sex, which were measured at the index date; all characteristics were included in the estimation of propensity scores and treated as estimators/factors in the decision tree analysis. [c]Severe hypoglycemia was only measured and included for propensity score estimation and the decision tree analysis in the secondary cohort. [d]Diabetes duration was measured as the time from the first date of type 2 diabetes diagnosis to the index date. [e]IS or TIA, other CVDs, and CVDs at baseline were measured mutually exclusively. [f]Severe hypoglycemia was defined as hospital admission or emergency department records for hypoglycemia.

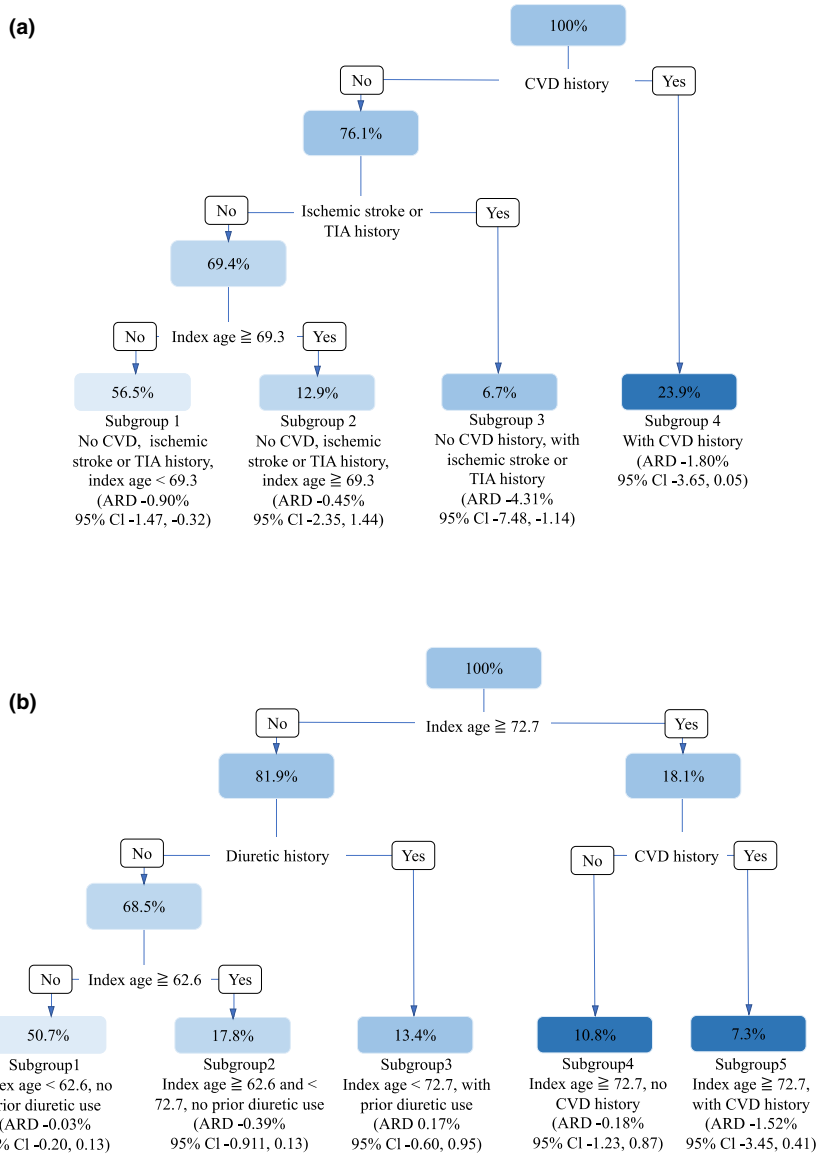

Figure 1 Subgroups classified by the tree analysis (analysis of primary cohort under intention-to-treat scenario). (**a**) Composite cardiovascular disease event as the outcome, (**b**) all-cause death as the outcome, and (**c**) three-point major adverse cardiovascular event as the outcome. Each node/subgroup specifies the percentage of subjects partitioned by the significant factors identified by the tree analysis. For example, in **a**, subgroup 1 included 56.6% of patients in the primary cohort who did not have cardiovascular disease (CVD), ischemic stroke (IS), or transient ischemic attack (TIA) history and were aged below 69.3 years when initiating treatments. The absolute risk difference (ARD) with its 95% confidence interval (CI) indicates the difference in the event rate of composite CVD between treatments. For example, subgroup 1 had a significant reduction in the ARD for composite CVD with dipeptidyl peptidase-4 inhibitor vs. sulfonylurea use of 0.90% (95% CI 0.32–1.47% decrease). [Colour figure can be viewed at wileyonlinelibrary.com]

the tree) determined/split by all significant factors selected in CART yielded maximum between-group differences and minimum within-group differences in treatment effects for a given study outcome. Furthermore, fivefold cross-validation was performed to ensure the validity of the estimators' performance.

For the most representative subgroups defined from the CART analysis, we further performed the Cox proportional hazard models to assess the hazards of study outcomes associated with treatment status in each subgroup. Results were presented as hazard ratios (HRs) and 95% confidence intervals (CIs). The stratified log-rank test was then performed to test the trend in the hazards of study outcomes across subgroups/nodes. Additionally, the absolute risk difference (ARD) in the event rates of study outcomes between treatments in each subgroup (node) was estimated. In contrast to conventional relative measures (e.g., HRs), ARD results provide a clinically meaningful absolute estimate,[32] and thus ARD is commonly used in studies of HTEs for assessing the performance of a decision tree analysis.[18,19] The c-statistic, which is equal to the area under the receiver operating characteristic curve, was reported to demonstrate the goodness of fit of the Cox models performed in each subgroup/node.[33] The significance of the goodness of fit for each Cox model was further assessed using the likelihood ratio test.

Additionally, to account for possible overestimation of treatment effects in the ITT scenario analyses, where nonadherence to medications was ignored, we performed analyses based on an as-treated (AT) scenario where patients who switched from or discontinued the study drug were also censored in addition to those censored points used in the ITT

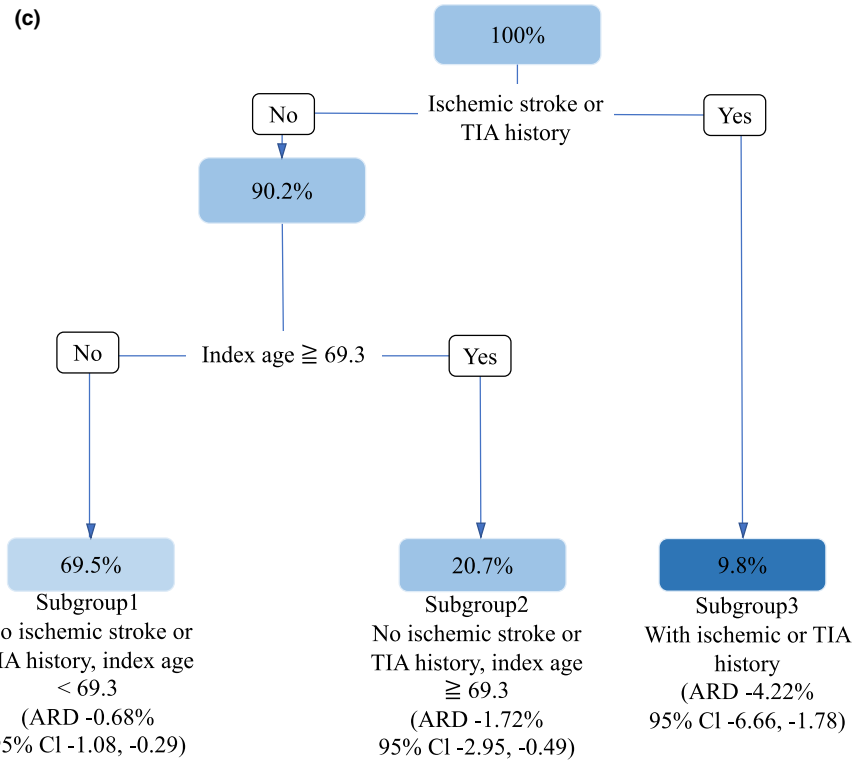

Figure 1 (Continued)

scenario analyses. Moreover, considering the potential link between hypoglycemia and CVD risks,[34] we constructed a secondary cohort where prior severe hypoglycemia (i.e., hospitalized hypoglycemia in the year prior to the index date) was adjusted in the analysis. Specifically, severe hypoglycemia was included as one of the variables for PS estimation in the beginning of deriving the study cohort and treated as a potential estimator/factor for splitting the data to determine the subgroups/nodes in the CART analysis. Tree analyses using the secondary cohort were performed under ITT and AT scenarios. A two-tail P value of < 0.05 was considered statistically significant. All analyses were conducted using R software version 3.6.1.

RESULTS

Table S3 presents the patient characteristics before the PS matching. Through the PS matching procedure, 19,853 and 19,856 PS-matched pairs of DPP-4i and SU stable users were included in the primary and secondary cohorts, respectively (**Table 1**). All patient characteristics were comparable between the treatment groups after the PS matching.

The overall analysis results based on the primary cohort under ITT scenario indicated that the use of DPP-4i vs. SU significantly reduced the risks of composite CVD and three-point MACE by 8% (HR 0.92, 95% CI 0.87–0.97) and 18% (HR 0.82, 95% CI 0.75–0.89), respectively, but had no significant effect for all-cause death (HR 0.94, 95% CI 0.80–1.10). **Figure 1a** shows CVD history, ischemic stroke (IS) or transient ischemic attack (TIA) history, and age at the index date, which split the study subjects into four subgroups/nodes with significantly different risks of composite CVD associated with DPP-4i vs. SU use, as reported in **Table 2**. Although the log-rank test results for the ARD in composite CVD between treatments are only statistically significant in subgroups 1 and 3 ($P < 0.05$), the stratified log-rank test for the difference in the hazards of composite CVD for DPP-4i vs. SU use across four subgroups is statistically significant ($P = 0.002$), which supports significant HTEs across these subgroups. Subgroup 1 (without CVD history, IS or TIA history, and age at the index date of < 69.3 years) comprised 56.5% of the sample and had a significant reduction in ARD for composite CVD with DPP-4i vs. SU use of 0.90% (95% CI 0.32–1.47% decrease, HR 0.86, 95% CI 0.77–0.97). Subgroup 3 (without CVD history but with IS or TIA history) accounted for 6.7% of the study subjects and had a significant reduction in ARDs for composite CVD with DPP-4i vs. SU use of 4.31% (95% CI 1.14–7.48% decrease, HR 0.81, 95% CI 0.69–0.95). The results of the c-statistic across the models in the 4 subgroups ranged from 0.646 (subgroup 3) to 0.719 (subgroup 1).

Figure 1b shows age at the index date, prior diuretic use, and CVD history, splitting the study subjects into five subgroups that had different mortality outcomes associated with DPP-4i vs. SU use. However, the heterogeneity of all-cause mortality with DPP-4i vs. SU use across these subgroups was not supported; the stratified log-rank test for the trend in the hazards of all-cause mortality across subgroups is not significant ($P = 0.442$; **Table 2**).

Figure 1c shows IS or TIA history and age at the index date, partitioning the study patients into three subgroups for different risks of three-point MACE between DPP-4i and SU use, as reported in **Table 2**. All log-rank tests for the ARD in three-point MACE between DPP-4i and SU use in individual subgroups

Table 2 Study outcomes associated with DPP-4i vs. SU use stratified by subgroups identified by the tree analysis (analysis of primary cohort under intention-to-treat scenario)

	No. of DPP-4i users (no. of events)	No. of SU users (no. of events)	ARD in event rate between treatments, % (95% CI)	*P* value of log-rank test for ARD	HR (95% CI)	C-statistic (95% CI)	*P* value of stratified log-rank test for difference in HRs across subgroups
Outcome: Composite CVD event							
Subgroup 1: No CVD, IS, or TIA history, index age < 69.3	11,311 (529)	11,123 (620)	−0.90 (−1.47, −0.32)	0.036	0.86 (0.77, 0.97)	0.719 (0.702, 0.737)	0.002
Subgroup 2: No CVD, IS, or TIA history, index age ≥ 69.3	2,597 (353)	2,506 (352)	−0.45 (−2.35, 1.44)	0.846	0.96 (0.83, 1.12)	0.663 (0.641, 0.685)	
Subgroup 3: No CVD history, with IS or TIA history	1,311 (271)	1,369 (342)	−4.31 (−7.48, −1.14)	0.027	0.81 (0.69, 0.95)	0.646 (0.623, 0.670)	
Subgroup 4: With CVD history	4,634 (1,354)	4,855 (1,506)	−1.80 (−3.65, 0.05)	0.324	0.99 (0.92, 1.06)	0.674 (0.663, 0.685)	
Outcome: All-cause death							
Subgroup 1: Index age < 62.6 years, no prior diuretic use	10,127 (35)	10,001 (38)	−0.03 (−0.20, 0.13)	0.981	1.07 (0.67, 1.69)	0.829 (0.760, 0.899)	0.442
Subgroup 2: Index age ≥ 62.6 and < 72.7 years, no prior diuretic use	3,453 (37)	3,621 (53)	−0.39 (−0.91, 0.13)	0.252	0.81 (0.53, 1.24)	0.811 (0.749, 0.874)	
Subgroup 3: Index age < 72.7 years, with prior diuretic use	2,675 (59)	2,658 (54)	0.17 (−0.60, 0.95)	0.410	1.08 (0.74, 1.56)	0.791 (0.735, 0.846)	
Subgroup 4: Index age ≥ 72.7 years, no CVD history	2,186 (67)	2,095 (68)	−0.18 (−1.23, 0.87)	0.865	0.97 (0.69, 1.36)	0.749 (0.698, 0.800)	
Subgroup 5: Index age ≥ 72.7 years, with CVD history	1,412 (97)	1,478 (124)	−1.52 (−3.45, 0.41)	0.302	0.86 (0.66, 1.13)	0.722 (0.682, 0.761)	
Outcome: 3-point MACE							
Subgroup 1: No ischemic stroke or TIA history, index age < 69.3 years	13,859 (356)	13,721 (446)	−0.68 (−1.08, −0.29)	0.010	0.81 (0.70, 0.93)	0.736 (0.715, 0.756)	< 0.001
Subgroup 2: No IS or TIA history, index age ≥ 69.3 years	4,142 (333)	4,088 (399)	−1.72 (−2.95, −0.49)	0.036	0.85 (0.73, 0.98)	0.667 (0.646, 0.689)	
Subgroup 3: With IS or TIA history	1,852 (306)	2,044 (424)	−4.22 (−6.66, −1.78)	0.004	0.80 (0.69, 0.93)	0.647 (0.625, 0.668)	

ARD, absolute risk difference; CI, confidence interval; CVD, cardiovascular disease; DPP-4i, dipeptidyl peptidase-4 inhibitor; HR, hazard ratio; IS, ischemic stroke; MACE, major adverse cardiovascular event; SU, sulfonylurea; TIA, transient ischemic attack.

and the stratified log-rank test for the difference in the hazards of three-point MACE across subgroups are statistically significant. Subgroup 1 (without IS or TIA history, and age at the index date of < 69.3 years), subgroup 2 (without IS or TIA history and age at the index date of ≥ 69.3 years), and subgroup 3 (with IS or TIA history) accounted for 69.5%, 20.7%, and 9.8% of the study subjects, respectively, and had significant reductions in the ARD for 3-point MACE with DPP-4i vs. SU use of 0.68% (95% CI 0.29–1.08% decrease, HR 0.81, 95% CI 0.70–0.93), 1.72% (95% CI 0.49–2.95% decrease, HR 0.85, 95% CI 0.73–0.98), and 4.22% (95% CI 1.78–6.66% decrease, HR 0.80, 95% CI 0.69–0.93), respectively. Among these subgroups, the c-statistic in the models ranged from 0.647 (subgroup 3) to 0.736 (subgroup 1).

Figure S2 shows the results of the analyses of the primary cohort under the AT scenario. The findings are generally consistent with those from the analyses under the ITT scenario (**Figure 1**), except for the all-cause death outcome; in the AT scenario, the tree analysis only partitioned the patients into three subgroups according to two variables, namely age at the index date and prior diuretic use (**Figure S2b**). Consistent with the results in **Table 2**, **Table S4** shows significant heterogeneity of composite CVD and three-point MACE between DPP-4i and SU use across the subgroups, as supported by statistically significant results of the stratified log-rank test for the trend in the hazards across subgroups ($P = 0.001$ and $P < 0.001$, respectively); the heterogeneity of all-cause mortality between treatments across the subgroups was not significant.

Figures S3 and **S4** show the results of the tree analyses for the secondary cohort (with consideration of severe hypoglycemia) under the ITT and AT scenarios, respectively. These results are consistent with the findings using the primary cohort for all three study outcomes regarding the factors/characteristics identified for subgroups, although the cutoff point for age at the index date for splitting the sample is slightly different between the analyses using the two different cohorts. **Tables S5** and **S6** show the detailed results of ARDs and HRs for study outcomes between treatments, statistical testing for the ARDs in individual subgroups, and the trend in the HRs across subgroups based on the secondary cohort, which are consistent with the findings from the analyses of the primary cohort (**Table 2** and **Table S4**), except the all-cause mortality outcome.

DISCUSSION

Existing studies commonly only report the average treatment effects between treatment groups. This study is the first of its kind to suggest significant HTEs on cardiovascular outcomes associated with the use of DPP-4i vs. SU across certain subgroups of patients with T2D. These heterogeneities may facilitate explanations of current inconsistent study results of comparative cardiovascular effects of DPP-4i vs. SU. We found that the relative risk of composite CVD for using DPP-4i vs. SU varied significantly in the four sets of patient subgroups stratified by a combination of three factors/characteristics: CVD history, IS or TIA history, and age at the date of initiating DPP-4i or SU. Among these subgroups, the patients with IS or TIA history but without any other established CVDs (subgroup 3) had the most risk reduction in the composite CVD by using DPP-4i instead of SU, followed by those without CVD, IS, or TIA history and initiated with DPP-4i at the age of < 69.3 years (subgroup 1). In addition, the relative risk of three-point MACE with using DPP-4i vs. SU varied in the three patient subgroups stratified by a combination of IS or TIA history and age at the date of initiating DPP-4i or SU; those with IS or TIA history (subgroup 3) had the most benefit of the reduced three-point MACE from using DPP-4i instead of SU, followed by those without IS or TIA history and with an age at the time of the initiating DPP-4i of < 69.3 years (subgroup 1) and those without IS or TIA history and with an age at the time of initiating DPP-4i of ≥ 69.3 years (subgroup 2). Moreover, although a possible link between hypoglycemia and CVD risks has been documented,[34] our separate analyses using primary and secondary cohorts yielded consistent results and the history of severe hypoglycemia was not selected as a significant factor in CART analyses using the secondary cohort, suggesting that relative to prior severe hypoglycemia, the CVD history, IS or TIA history, and age of initiating drugs may play a more important role in explaining the comparative CVD risks of using DPP-4i vs. SU. Because we considered severe hypoglycemia before treatment initiation, one could argue that prior severe hypoglycemia episodes might have a limited impact on the future development of CVDs.

Although no comparable studies have assessed the HTEs of DPP-4i vs. SU on CVD risks, numerous studies have used the conventional univariable analysis to characterize subgroups that may vary in treatment effects. However, most studies analyzed treatment effects only in a single subgroup of patients with established CVD events but did not assess other subgroups, such as patients without CVD history.[8,9,11,12,15,16,35,36] Such analyses cannot answer whether there is an interaction between the status of CVD history (yes vs. no) and treatment status (DPP-4i vs. SU). Furthermore, in the patient subgroup with established CVD events, the results of comparative CVD risks for DPP-4i vs. SU are inconsistent; some studies showed that the use of DPP-4i vs. SU yielded a significantly lower CVD risk,[9,11,12,17,36] whereas others demonstrated no difference between DPP-4i and SU.[8,35] Demographic variables, such as age at treatment initiation[37,38] and sex,[13,37–39] were also considered in the previous univariable analyses. However, the results of these demographic subgroups were inconsistent[13,37–39] and the cutoff point for the stratification of age subgroups varied among studies.[37,38]

Considering that heterogeneous patient subgroups that vary in treatment responses in real-world clinical practice cannot be easily characterized by a single demographic or clinical factor/characteristic, the complexity of the relationships between multiple factors and treatment outcomes is unlikely to be revealed by conventional univariable subgroup analyses. Real-world patient subgroups are likely to have mixed factors, the interactions of which could influence treatment responses. Hence, it has been argued that conventional univariable analyses lack the statistical power to detect heterogeneity of treatment effects among subgroup patients.[18,19] Univariable analysis results should be cautiously applied because of potential false-positive biases due to multiple testing errors (e.g., overfitting due to multiple comparisons).[32,40,41]

Compared with previous studies on cardiovascular safety of DPP-4i vs. SU, the present study has several methodological strengths. First, in contrast to previous univariable analyses, we used advanced machine-learning approaches to consider multiple simultaneous covariates and the interactions among them, which would thus provide better estimations and explanations for variations in the treatment effects. Second, although some factors associated with the HTEs of DPP-4i vs. SU identified in this study (i.e., CVD history and age) are consistent with the findings from previous univariable analyses, the assessment of patient subsets stratified by the combination of multiple factors in this study not only detected the significant treatment effects within each patient subset but also examined the significant HTEs across different patient subsets. Third, for a continuous variable (e.g., age), the data-driven machine-learning approach determined the cutoff point based on the best data fit for study subjects, which avoided arbitrary assumptions or assignments on study variables. Fourth, the machine-learning analysis in this study avoided the overfitting issue commonly seen in traditional univariable analyses, which led to unbiased prediction for the differences in treatment effects between treatment groups. From a clinical perspective, there are several strengths and implications of this study. First, compared with the results of conventional univariable subgroup analyses, our findings based on data-driven multivariable subgroup analyses are more clinically relevant to real-world patient populations, which are commonly classified by multiple factors at the same time, and thus can assist clinician-patient discussions about selecting GLAs. Second, using rigorous machine-learning analytic methods with cross-validation procedures, it is possible to identify subgroups of patients for which there is a substantial difference in treatment effects (i.e., beneficial or detrimental). Third, considering that multiple factors in combination may be required to explain clinically important variations in benefit or harm seen in real-world settings, the treatment responses (i.e., ARD estimates presented in this study) of patient subgroups are more readily interpretable by clinicians and patients and can thus facilitate personalized treatment decisions. Last, the visualized results of a decision tree analysis can be used to prioritize a series of important clinical factors/characteristics to facilitate personalized treatment decisions.

This study has several potential limitations. First, the causal inferences from study results should be cautiously applied and further verified in future studies because of the hypothesis-generating nature in this exploratory research. Particularly, future prospective studies are needed to determine complex physiological relationships among the selected variables for the stratification of patient subsets, which have not been proven for causality in our variable selection procedures. Second, residual confounding by unmeasured factors (e.g., prescribers' preferences or behaviors, patients' preferences or their lifestyle behaviors, and laboratory data) may not be ruled out in this claims-based study. However, efforts to minimize unmeasured confounding have been made through the matching procedure that achieved a greater level of between-group comparability on various baseline patient characteristics. Third, because the two subgroups we tested were defined by the same covariates, the two tests for heterogeneity might not be independent.[19] Fourth, from the perspective of clinical decisions, we only analyzed the baseline patient characteristics before or at treatment initiation as the potential factors of HTEs, rather than considered them as time-varying variables over time. Time-varying changes or variables measured after treatment initiation and before study outcome occurrence may also be associated with treatment outcomes. Fifth, the PS matching procedure enhanced the internal validity of the study, but the external validity may be concerned because a certain proportion of patients was excluded after the matching process and all study subjects were Taiwanese patients. Hence, future research utilizing external databases comprised of other ethnic populations is warranted to assess the external validity of our findings. Last, the small number of death events may limit the power of this study to identify significant heterogeneous subgroups for this outcome.

In summary, this exploratory study suggests significant HTEs of using DPP-4i vs. SU on cardiovascular outcomes in patients with T2D. The factors attributable to possible HTEs include the CVD history, IS or TIA history, and age at drug initiation. Our analyses suggest that the use of DPP-4i instead of SU might be prioritized for patients with prior IS or TIA and for those aged 69 years or lower without prior CVDs, and IS or TIA. More broadly, this study highlights the importance of exploring clinically meaningful relationships among treatment interventions, clinical outcomes, and patient characteristics to enhance personalized healthcare and efficient healthcare resource allocation to patients in need. Future research is encouraged to apply our analytical framework to explore potential HTEs of newer classes of antidiabetic drugs (i.e., GLP-1 receptor agonists and SGLT2 inhibitors) to guide individualized antidiabetic treatment and improve the quality of diabetes care.

SUPPORTING INFORMATION

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

ACKNOWLEDGMENT

Technical/expert advice for this project was provided by the Michigan Center for Diabetes Translational Research (grant P30DK092926 from the National Institute of Diabetes and Digestive and Kidney Diseases) (Shihchen Kuo).

FUNDING

This project was supported by a grant from the Ministry of Science and Technology in Taiwan (grant MOST 109-2320-B-006-047-MY3) (H.-T.O.). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

CONFLICTS OF INTEREST

All authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

C.Y.Y., H.T.O., and S.K. wrote the manuscript. C.Y.Y., H.T.O., and S.K. designed the research. C.Y.Y., H.T.O., W.A.L., P.F.S., L.J.L., and C.T.Y. performed the research. C.Y.Y., W.A.L., and P.F.S. analyzed the data. W.A.L. and P.F.S. contributed new reagents/analytical tools.

1. Ou, H.T., Chang, K.C., Liu, Y.M. & Wu, J.S. Recent trends in the use of antidiabetic medications from 2008 to 2013: a nation-wide population-based study from Taiwan. *J. Diabetes* **9**, 256–266 (2017).
2. American Diabetes Association. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2020. *Diabetes Care* **43**, S98–S110 (2020).
3. Green, J.B. *et al*. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. *N. Engl. J. Med.* **373**, 232–242 (2015).
4. Scirica, B.M. *et al*. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N. Engl. J. Med.* **369**, 1317–1326 (2013).
5. White, W.B. *et al*. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. *N. Engl. J. Med.* **369**, 1327–1335 (2013).
6. 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).
7. 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).
8. Toh, S. *et al*. Prospective postmarketing surveillance of acute myocardial infarction in new users of saxagliptin: a population-based study. *Diabetes Care* **41**, 39–48 (2018).
9. O'Brien, M.J. *et al*. Association of second-line antidiabetic medications with cardiovascular events among insured adults with type 2 diabetes. *JAMA Netw. Open* **1**, e186125 (2018).
10. Cho, Y.Y. & Cho, S.I. Metformin combined with dipeptidyl peptidase-4 inhibitors or metformin combined with sulfonylureas in patients with type 2 diabetes: a real world analysis of the South Korean national cohort. *Metabolism* **85**, 14–22 (2018).
11. Ou, H.T., Chang, K.C., Li, C.Y. & Wu, J.S. Comparative cardiovascular risks of dipeptidyl peptidase 4 inhibitors with other second- and third-line antidiabetic drugs in patients with type 2 diabetes. *Br. J. Clin. Pharmacol.* **83**, 1556–1570 (2017).
12. Kim, Y.G. *et al*. Dipeptidyl peptidase-4 inhibitors and risk of heart failure in patients with type 2 diabetes mellitus: a population-based cohort study. *Circ. Heart Fail.* **10**, e003957 (2017).
13. Ha, K.H., Kim, B., Choi, H., Kim, D.J. & Kim, H.C. Cardiovascular events associated with second-line anti-diabetes treatments: analysis of real-world Korean data. *Diabet. Med.* **34**, 1235–1243 (2017).
14. Gordon, J., McEwan, P., Evans, M., Puelles, J. & Sinclair, A. Managing glycaemia in older people with type 2 diabetes: a retrospective, primary care-based cohort study, with economic assessment of patient outcomes. *Diabetes Obes. Metab.* **19**, 644–653 (2017).
15. Gokhale, M. *et al*. No increased risk of cardiovascular events in older adults initiating dipeptidyl peptidase-4 inhibitors vs therapeutic alternatives. *Diabetes Obes. Metab.* **19**, 970–978 (2017).
16. Toh, S. *et al*. Risk for hospitalized heart failure among new users of saxagliptin, sitagliptin, and other antihyperglycemic drugs: a retrospective cohort study. *Ann. Intern. Med.* **164**, 705–714 (2016).
17. Fu, A.Z. *et al*. Association between hospitalization for heart failure and dipeptidyl peptidase 4 inhibitors in patients with type 2 diabetes: an observational study. *Diabetes Care* **39**, 726–734 (2016).
18. Basu, S., Raghavan, S., Wexler, D.J. & Berkowitz, S.A. Characteristics associated with decreased or increased mortality risk from glycemic therapy among patients with type 2 diabetes and high cardiovascular risk: machine learning analysis of the ACCORD trial. *Diabetes Care* **41**, 604–612 (2018).
19. Baum, A., Scarpa, J., Bruzelius, E., Tamler, R., Basu, S. & Faghmous, J. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial. *Lancet Diabetes Endocrinol.* **5**, 808–815 (2017).
20. Hsieh, C.Y. *et al*. Taiwan's National Health Insurance Research Database: past and future. *Clin. Epidemiol.* **11**, 349–358 (2019).
21. Cheng, T.M. Taiwan's new national health insurance program: genesis and experience so far. *Health Aff. (Millwood)* **22**, 61–76 (2003).
22. Jiang, Y.D., Chang, C.H., Tai, T.Y., Chen, J.F. & Chuang, L.M. Incidence and prevalence rates of diabetes mellitus in Taiwan: analysis of the 2000–2009 Nationwide Health Insurance database. *J. Formos Med. Assoc.* **111**, 599–604 (2012).
23. Sheen, Y.J., Hsu, C.C., Jiang, Y.D., Huang, C.N., Liu, J.S. & Sheu, W.H. Trends in prevalence and incidence of diabetes mellitus from 2005 to 2014 in Taiwan. *J. Formos. Med. Assoc.* **118** (suppl. 2), S66–S73 (2019).
24. Ou, H.T., Chang, K.C., Li, C.Y. & Wu, J.S. Risks of cardiovascular diseases associated with dipeptidyl peptidase-4 inhibitors and other antidiabetic drugs in patients with type 2 diabetes: a nation-wide longitudinal study. *Cardiovasc. Diabetol.* **15**, 41 (2016).
25. Cheng, C.L., Kao, Y.H., Lin, S.J., Lee, C.H. & Lai, M.L. Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol. Drug Saf.* **20**, 236–242 (2011).
26. Cheng, C.L., Lee, C.H., Chen, P.S., Li, Y.H., Lin, S.J. & Yang, Y.H. Validation of acute myocardial infarction cases in the national health insurance research database in Taiwan. *J. Epidemiol.* **24**, 500–507 (2014).
27. Kuo, S., Yang, C.T., Wu, J.S. & Ou, H.T. Effects on clinical outcomes of intensifying triple oral antidiabetic drug (OAD) therapy by initiating insulin versus enhancing OAD therapy in patients with type 2 diabetes: a nationwide population-based, propensity-score-matched cohort study. *Diabetes Obes. Metab.* **21**, 312–320 (2019).
28. Cheng, C.L., Chien, H.C., Lee, C.H., Lin, S.J. & Yang, Y.H. Validity of in-hospital mortality data among patients with acute myocardial infarction or stroke in National Health Insurance Research Database in Taiwan. *Int. J. Cardiol.* **201**, 96–101 (2015).
29. Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. *Classification and Regression Trees.* 151–166 (Wadsworth. International Group 432, Belmont, CA, 1984).
30. De'ath, G. & Fabricius, K.E. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178–3192 (2000).
31. LeBlanc, M. & Crowley, J. Relative risk trees for censored survival data. *Biometrics* **48**, 411–425 (1992).
32. VanderWeele, T.J. & Knol, M.J. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann. Intern. Med.* **154**, 680–683 (2011).
33. Harrell, J. & Frank, E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (Springer, Cham, 2015).
34. Yun, J.S., Park, Y.M., Han, K., Cha, S.A., Ahn, Y.B. & Ko, S.H. Severe hypoglycemia and the risk of cardiovascular disease and mortality in type 2 diabetes: a nationwide population-based cohort study. *Cardiovasc. Diabetol.* **18**, 103 (2019).
35. Kim, K.J. *et al*. Dipeptidyl peptidase-4 inhibitor compared with sulfonylurea in combination with metformin: cardiovascular and renal outcomes in a propensity-matched cohort study. *Cardiovasc. Diabetol.* **18**, 28 (2019).
36. Chin, H.J., Nam, J.H., Lee, E.K. & Shin, J.Y. Comparative safety for cardiovascular outcomes of DPP-4 inhibitors versus glimepiride in patients with type 2 diabetes: a retrospective cohort study. *Medicine (Baltimore)* **96**, e7213 (2017).
37. Ou, S.M. *et al*. Effects on clinical outcomes of adding dipeptidyl peptidase-4 inhibitors versus sulfonylureas to metformin therapy in patients with type 2 diabetes mellitus. *Ann. Intern. Med.* **163**, 663–672 (2015).

38. Morgan, C.L., Mukherjee, J., Jenkins-Jones, S., Holden, S.E. & Currie, C.J. Combination therapy with metformin plus sulphonylureas versus metformin plus DPP-4 inhibitors: association with major adverse cardiovascular events and all-cause mortality. *Diabetes Obes. Metab.* **16**, 977–983 (2014).
39. Seong, J.M. *et al*. Differential cardiovascular outcomes after dipeptidyl peptidase-4 inhibitor, sulfonylurea, and pioglitazone therapy, all in combination with metformin, for type 2 diabetes: a population-based cohort study. *PLoS One* **10**, e0124287 (2015).
40. Burke, J.F., Hayward, R.A., Nelson, J.P. & Kent, D.M. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ. Cardiovasc. Qual. Outcomes* **7**, 163–169 (2014).
41. Hayward, R.A., Kent, D.M., Vijan, S. & Hofer, T.P. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med. Res. Methodol.* **6**, 18 (2006).