Heterogeneity of Treatment Effects Across Nine Glucose-Lowering Drug Classes in Type 2 Diabetes: Extension of the LEGEND-T2DM Network Study

Hsin Yi Chen BS[1], Thomas Falconer MS[1], Anna Ostropolets[1,2] MD, PhD, Tara V. Anand MA, MPhil[1], Xinzhuo Jiang[1] MS, David Dávila-García MA[3], Linying Zhang PhD[3], Ruochong Fan MA[3], Hannah Morgan-Cooper[4] MSci, George Hripcsak MD, MS[1]

[1] Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY
[2]Janssen Research & Development, Titusville, NJ
[3] Institute for Informatics, Data Science and Biostatistics, Washington University in St. Louis, St. Louis, MO
[4] Stanford School of Medicine and Stanford Health Care, Palo Alto, CA

Background

Type 2 diabetes mellitus (T2DM) is a major cause of morbidity and mortality, affecting more than 525 million people globally[1]. Therapeutic options for T2DM have expanded over the last decade with the emergence of sodium-glucose co-transporter-2 inhibitors (SGLT2i) and glucagon-like peptide-1 receptor agonists (GLP-1RA), which reduced the risk of major cardiovascular events in randomized controlled trials[2–5]. OHDSI's LEGEND-T2DM study[6,7] filled our gap in understanding of the relative effects of T2DM agents on cardiovascular risk and patient-centered safety outcomes by conducting direct head-to-head comparisons of second-line antihyperglycemic drugs. However, patients with T2DM are a heterogeneous group, varying widely in terms of age, sex, race, and comorbidities[8]. These factors may significantly modify the benefits and risks associated with different antihyperglycemic drugs[9–11]. This present observational study extend OHDSI's original LEGEND-T2DM study by systematically investigating the heterogeneity of treatment effects for nine major classes of antihyperglycemic drugs across a comprehensive array of predefined clinical and demographic subgroups. Specifically, we aim to determine if the comparative effectiveness and safety of these drug classes, for outcomes including major adverse cardiovascular events, renal events, and other patient-centered safety endpoints, differ significantly based on patient characteristics such as demographics, diabetes severity, comorbidities, and other relevant clinical history.

Methods

We conducted a large-scale, multinational, real-world comparative effectiveness and safety study, extending the LEGEND-T2DM[6,7] and LEGEND-HTN studies[12].

The study population comprised adults (≥18 years of age) diagnosed with T2DM who initiated treatment with a drug agent from one of the nine specified glucose-lowering drug classes: (1) Alpha-Glucosidase Inhibitors, (2) Biguanides, (3) Dipeptidyl Peptidase-4 Inhibitors (DPP-4i), (4) dual Glucose-dependent Insulinotropic Polypeptide (GIP) and Glucagon-Like Peptide-1 Receptor Agonists (GLP-1RA), (5) GLP-1RA, (6) Meglitinides, (7) Sodium-Glucose Cotransporter-2 Inhibitors (SGLT2-i), (8) Sulfonylureas (SU), and (9) Thiazolidinediones. To assess for heterogeneity of treatment effects (HTE), we stratified patients by predefined clinical and demographic subgroups at baseline (Table 1). We evaluated a total of 11 study outcomes: acute myocardial infarction, acute renal failure, hospitalization for heart failure, stroke, abnormal weight gain, acute pancreatitis, diabetic ketoacidosis, diarrhea, hypoglycemia, vomiting, and

hepatic failure.

Pairwise comparisons were performed using calibrated hazard ratios (HRs) for each target-comparator-outcome-subgroup combination. The HR represents the relative hazard, h(t), of experiencing an outcome at time t between individuals receiving the target vs. comparator drug classes (Equation 1).

$$HR = \frac{h_{\text{target}}(t)}{h_{\text{comparator}}(t)} \quad 1$$

HTE was quantified by calculating the difference in log transformed HRs, $\Delta \ln(HR)$, between subgroups (e.g., Male vs. Female; Age <21 vs. Age 21-60; Hypertension vs. No Hypertension) within the same drug-outcome comparison (Equation 2). The corresponding standard error associated with this difference ($SE_{\Delta \ln(HR)}$) was computed using the subgroup-specific SEs (Equation 3).

$$\Delta \ln(HR) = \ln\left(HR_{\text{subgroup1}}\right) - \ln\left(HR_{\text{subgroup2}}\right) \quad 2$$

$$SE_{\Delta \ln(HR)} = \sqrt{\left(SE^2_{\ln(HR_{\text{subgroup1}})} + SE^2_{\ln(HR_{\text{subgroup2}})}\right)} \quad 3$$

$\Delta \ln(HR)$ and $SE_{\Delta \ln(HR)}$ were then pooled across databases using random-effects meta-analysis. While we calculated Z-scores (Equation 4) and their corresponding p-values, we focused on ranking the five lowest ("most significant") p-values within each subgroup rather than applying significance thresholds, as this was a hypothesis-generating study.

$$Z = \frac{\Delta \ln(HR)}{SE_{\Delta \ln(HR)}} \quad 4$$

The detailed study protocol, including concept sets for our cohorts, can be found at https://github.com/ohdsi-studies/LegendT2dmArpah.

Subgroup	Comparisons
Age	Pairwise comparisons between: • Younger: <21 years • Middle-aged: 21–60 years • Older: >60 years
Biological Sex	Comparison between Female and Male (as recorded in database based on SNOMED codes)
Renal Impairment	Stratified into three categories based on diagnosis codes for chronic kidney disease (CKD) and end-stage renal disease (ESRD), dialysis procedures, and relevant laboratory measures (e.g., estimated glomerular filtration rate (eGFR), serum creatinine, urine albumin). We compare: • No renal impairment (NRI) vs Renal impairment without dialysis (RI-ND) • No renal impairment (NRI) vs Renal impairment on dialysis (RI-D)
Obesity	Presence of obesity, defined as a Body Mass Index (BMI) >30 kg/m², body weight >120 kg, or a diagnosis code for obesity. Compared against the non-obese subgroup.
Poorly Controlled Diabetes	Defined as an HbA1c >8% (64 mmol/mol) or a diagnosis code indicating uncontrolled type 2 diabetes or poor diabetes control. Compared against those not meeting these criteria.

Diabetic Ketoacidosis (DKA)	History of DKA based on diagnosis codes. Compared against those with no history of DKA.
Diabetic Retinopathy	History of diabetic retinopathy based on diagnosis codes. Compared against those with no history of diabetic retinopathy.
Essential Hypertension	History of essential hypertension based on diagnosis codes. Compared against those with no history of essential hypertension.
Hyperlipidemia	History of hyperlipidemia based on diagnosis codes. Compared against those with no history of hyperlipidemia.
Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD)	History of MASLD (including non-alcoholic fatty liver disease or non-alcoholic steatohepatitis) based on diagnosis codes. Compared against those with no history of MASLD.

Table 1. Subgroup comparisons in this study.

Results

We identified the top five pairwise drug comparisons with the lowest p-values in each subgroup as potential signals of HTE (Table 2).

Our study was run on a total of 6 databases: Columbia University Irving Medical Center (CUIMC), Washington University of St. Louis (WashU), Japan Medical Data Center (JMDC), Merative MarketScan Multi-State Medicaid Database (MDCD), Merative MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR), and Stanford Healthcare (STARR OMOP). Data from WashU was excluded due to failed diagnostics. Several subgroup comparisons were also excluded due to failed diagnostics, including age (younger vs. older, younger vs. middle-aged), renal impairment (all comparisons), poorly controlled diabetes, diabetic ketoacidosis, diabetic retinopathy, and MASLD.

Clinical subgroups. Biguanides had a stronger protective effect (compared to DPP-4i) against acute myocardial infarctions in the hyperlipidemia (HLD) group, and while there were no definitive benefits or risks for taking biguanide vs. SU for abnormal weight gain, or for taking biguanide vs. SGLT-2i for stroke, our results show that there may exist a differential effect based on lipid status. In the hypertension subgroup, there was no difference in risk of acute myocardial infarction between biguanide and SU, but our results show that there may be a potential benefit of biguanides over SU in the non-hypertensive subgroup. Finally, for the obesity subgroup, biguanide had a protective effect (compared to DPP-4i) against hospitalization with heart failure for obese patients, while SGLT-2i (compared to GLP-1 RA) had a protective effect against stroke only in the non-obese group. Finally, non-obese patients taking SU (vs. GLP-1 RA) had a higher risk of hypoglycemia than obese patients taking SU (vs. GLP-1 RA).

Demographic subgroups. Female patients on GLP-1 RA (vs. DPP-4i) were more likely to experience diarrhea, and female patients on SGLT-2i (vs. DPP-4i) were more protected against stroke. These effects were not seen for male patients. Age-based comparisons did not meet typical significance thresholds but included differences for diabetic ketoacidosis (biguanide vs. GLP-1 RA and SU vs. SGLT-2i) and hypoglycemia (SU vs. GLP-1 RA).

Our findings aligned with known pharmacologic patterns in the literature, such as the weight-loss benefits of metformin[13], the cardioprotective benefits of GLP-1 RA[5,14], and how gastrointestinal side

effects are common for those taking GLP-1 RA[15–17]. Additionally, our study identified potential areas subgroup heterogeneity, such as a lower risk of heart failure hospitalization with biguanide (vs. DPP-4i) or a lower risk of stroke with GLP-1 RA (vs. SGLT-2i) for obese patients. A potential explanation for this is the weight-loss benefit of metformin relative to DPP-4i (which is weight neutral), and GLP-1 RA vs. SGLT-2i (modest weight loss, but less so than GLP-1 RA[18,19]): because higher BMI is associated with higher rates of cardiovascular events, the absolute benefit of an anti-hyperglycemia drug that also promotes weight loss might be larger in obese individuals. Similarly, our results showed that female patients experienced diarrhea more often on GLP-1 RA (vs. DPP-4i), consistent with literature showing that women experience GI side effects with GLP-1 RAs at roughly twice the rate of men[20–22], which could be attributed to gender differences in adverse event reporting[23].

Hyperlipidemia (HLD) Subgroups (HLD vs. No HLD)					
Outcome	**Target**	**Comparator**	**HR (HLD)**	**HR (No HLD)**	**p-value**
Abnormal weight gain	Biguanide	SU	2.97 (0.92,9.56)	0.45 (0.18,1.14)	**0.01**
Stroke	Biguanide	SGLT-2i	1.76 (0.91,3.43)	0.73 (0.44,1.23)	**0.04**
Acute myocardial infarction	Biguanide	DPP-4i	0.55 (0.35,0.85)†	1.13 (0.65,1.98)	**0.04**
Acute pancreatitis	Biguanide	DPP-4i	0.74 (0.43,1.27)	1.73 (0.89,3.36)	0.05
Diarrhea	Biguanide	DPP-4i	0.97 (0.67,1.4)	1.75 (1.08,2.82)	0.06
Hypertension (HTN) Subgroups (HTN vs. No HTN)					
Outcome	**Target**	**Comparator**	**HR (HTN)**	**HR (No HTN)**	**p-value**
Acute myocardial infarction	Biguanide	SU	1.65 (0.92, 2.95)	0.78 (0.51,1.17)	**0.04**
Hospitalization with heart failure events	SU*	DPP-4i*	0.84 (0.71, 1)	1.07 (0.9, 1.28)	0.05
Hypoglycemia	Biguanide	DPP-4i	0.89 (0.53, 1.5)	0.45 (0.24, 0.84)†	0.11
Hypoglycemia	Biguanide	SGLT-2i	1.37 (0.67, 2.79)	0.54 (0.21,1.37)	0.12
Acute myocardial infarction	Biguanide	DPP-4i	0.51 (0.27, 0.95)†	0.9 (0.6, 1.35)	0.14
Obesity Subgroups (Obese vs. Non-Obese)					
Outcome	**Target**	**Comparator**	**HR (Obese)**	**HR (Non-Obese)**	**p-value**
Hospitalization with heart failure events	Biguanide	DPP-4i	0.53 (0.37, 0.75)†	0.93 (0.76,1.15)	**0.01**
Stroke	SGLT-2i*	GLP-1 RA*	1.5 (0.82, 2.72)	0.51 (0.26, 0.99)†	**0.02**
Hypoglycemia	SU*	GLP-1 RA*	1.21 (0.86, 1.7)	2.35 (1.39, 3.97)†	**0.03**
Acute pancreatitis	Biguanide	SU	1.96 (0.99, 3.88)	0.82 (0.46, 1.47)	0.06
Hepatic failure	Biguanide	SU	1.54 (0.73, 3.27)	0.61 (0.32, 1.16)	0.07
Sex Subgroups (Male vs. Female)					
Outcome	**Target**	**Comparator**	**HR (Male)**	**HR (Female)**	**p-value**
Diarrhea	GLP-1 RA*	DPP-4i*	0.8 (0.49,1.28)	1.43 (1.1, 1.84)†	**0.03**
Stroke	SGLT-2i*	DPP-4i*	0.95 (0.68,1.32)	0.38 (0.17, 0.86)†	**0.04**
Diarrhea	SU*	DPP-4i*	0.84 (0.66,1.07)	1.19 (0.95,1.5)	**0.04**
Diabetic ketoacidosis	GLP-1 RA*	DPP-4i*	1.76 (0.77,4.05)	0.66 (0.37,1.17)	0.06
Acute pancreatitis	SGLT-2i *	DPP-4i*	1.28 (0.83,1.98)	0.55 (0.21,1.42)	0.11
Age Subgroups (Older (>60 years) vs. Middle-aged (21-60 years))					
Outcome	**Target**	**Comparator**	**HR (>60y)**	**HR (21-60y)**	**p-value**
Diabetic ketoacidosis	Biguanide	GLP-1 RA	0.73 (0.22, 2.4)	2.84 (0.82, 9.8)	0.12
Hypoglycemia	SU*	GLP-1 RA*	2.84 (1.51,5.33)†	1.41 (0.74, 2.67)	0.12
Diabetic ketoacidosis	SU*	SGLT-2i *	0.22 (0.06, 0.79)	0.69 (0.34,1.37)	0.12
Abnormal weight gain	Biguanide	GLP-1 RA	1.16 (0.44, 3.09)	0.35 (0.11,1.17)	0.13
Acute renal failure	Biguanide	SGLT-2i	2.15 (0.62,7.45)	0.7 (0.29, 1.69)	0.15

Table 2. Results for our HTE analysis. Hazard ratio estimates (HRs) and 95% confidence intervals from individual data

sources were aggregated using random effects meta-analysis, followed by calibration using empirical null distributions to correct for residual confounding and systematic bias. The reported p-value corresponds to the statistical test of the difference between these subgroup-specific log hazard ratios, $\Delta\ln(HR)$.

* cohort definition for drug exposure did not explicitly exclude metformin use
† HR confidence interval does *not* include 1 and thus significantly favors one of the drug comparators
Abbreviations: DPP-4i: dipeptidyl peptidase-4 inhibitor(s), GLP-1 RA: glucagon-like peptide-1 receptor agonist(s), HR: hazard ratio, SGLT-2i: sodium-glucose cotransporter 2 inhibitor(s), SU: sulfonylureas

Conclusion

This hypothesis-generating study identified several potential signals where there exists treatment effect heterogeneity for several classes of T2DM drugs. While many findings did not meet the significance threshold, and this study only included a few databases, our preliminary findings highlight the potential for personalized T2DM treatment recommendations based on patient characteristics.

References

1. 2024 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association [Internet]. [cited 2025 May 14]. Available from: https://www.ahajournals.org/doi/epub/10.1161/CIR.0000000000001209

2. Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, et al. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. N Engl J Med. 2022 Sept 22;387(12):1089–98.

3. Hernandez AF, Green JB, Janmohamed S, D'Agostino RB, Granger CB, Jones NP, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. Lancet. 2018 Oct 27;392(10157):1519–29.

4. Andrikou E, Tsioufis C, Andrikou I, Leontsinis I, Tousoulis D, Papanas N. GLP-1 receptor agonists and cardiovascular outcome trials: An update. Hellenic J Cardiol. 2019;60(6):347–51.

5. Pablo A, Evelyn B, Claudia F, Yanina MA. GLP-1RA and SGLT2i: Cardiovascular Impact on Diabetic Patients. Curr Hypertens Rev. 2021;17(2):149–58.

6. Khera R, Schuemie MJ, Lu Y, Ostropolets A, Chen R, Hripcsak G, et al. Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies. BMJ Open. 2022 June 1;12(6):e057977.

7. Rohan Khera MD, Arya Aminorroaya MD, Lovedeep Singh Dhingra M, Phyllis M. Thangaraj MD, Aline Pedroso Camargos P, Fan Bu P, et al. Comparative Effectiveness of Second-Line Antihyperglycemic Agents for Cardiovascular Outcomes: A Multinational, Federated Analysis of LEGEND-T2DM. Journal of the American College of Cardiology [Internet]. 2024 Sept 3 [cited 2025 May 12]; Available from: https://www.jacc.org/doi/10.1016/j.jacc.2024.05.069

8. Trischitta V, Prudente S, Doria A. Disentangling the heterogeneity of adulthood-onset non-autoimmune diabetes: a little closer but lot more to do. Current Opinion in Pharmacology. 2020 Dec 1;55:157–64.

9. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. The Lancet Diabetes & Endocrinology. 2018 May 1;6(5):361–9.

10. Nair ATN, Wesolowska-Andersen A, Brorsson C, Rajendrakumar AL, Hapca S, Gan S, et al. Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. Nat Med. 2022 May;28(5):982–8.

11. Redondo MJ, Hagopian WA, Oram R, Steck AK, Vehik K, Weedon M, et al. The clinical consequences of heterogeneity within and between different diabetes types. Diabetologia. 2020 Oct 1;63(10):2040–8.

12. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. The Lancet. 2019 Nov;394(10211):1816–26.

13. Apovian CM, Okemah J, O'Neil PM. Body Weight Considerations in the Management of Type 2 Diabetes. Adv Ther. 2019;36(1):44–58.

14. Ferhatbegović L, Mršić D, Macić-Džanković A. The benefits of GLP1 receptors in cardiovascular diseases. Front Clin Diabetes Healthc. 2023 Dec 8;4:1293926.

15. Nauck MA, Meier JJ. Are all GLP-1 agonists equal in the treatment of type 2 diabetes? Eur J Endocrinol. 2019 Dec;181(6):R211–34.

16. Nauck MA, Quast DR, Wefers J, Meier JJ. GLP-1 receptor agonists in the treatment of type 2 diabetes – state-of-the-art. Molecular Metabolism. 2021 Apr 1;46:101102.

17. Bettge K, Kahle M, Abd El Aziz MS, Meier JJ, Nauck MA. Occurrence of nausea, vomiting and diarrhoea reported as adverse events in clinical trials studying glucagon-like peptide-1 receptor agonists: A systematic analysis of published clinical trials. Diabetes, Obesity and Metabolism. 2017 Mar;19(3):336–47.

18. Lingvay I, Catarig AM, Frias JP, Kumar H, Lausvig NL, le Roux CW, et al. Efficacy and safety of once-weekly semaglutide versus daily canagliflozin as add-on to metformin in patients with type 2 diabetes (SUSTAIN 8): a double-blind, phase 3b, randomised controlled trial. Lancet Diabetes Endocrinol. 2019 Nov;7(11):834–44.

19. Ma H, Lin YH, Dai LZ, Lin CS, Huang Y, Liu SY. Efficacy and safety of GLP-1 receptor agonists versus SGLT-2 inhibitors in overweight/obese patients with or without diabetes mellitus: a systematic review and network meta-analysis. BMJ Open. 2023 Mar 7;13(3):e061807.

20. Joung KI, Jung GW, Park HH, Lee H, Park SH, Shin JY. Gender differences in adverse event reports

associated with antidiabetic drugs. Sci Rep. 2020 Oct 16;10(1):17545.

21. de Vries ST, Denig P, Ekhart C, Mol PGM, van Puijenbroek EP. Sex Differences in Adverse Drug Reactions of Metformin: A Longitudinal Survey Study. Drug Saf. 2020 May 1;43(5):489–95.

22. Roseberry T, Grossrubatscher I, Krausz T, Wang Y, Schwartz M, Tingley D. Sex differences in GLP-1 signaling across species [Internet]. bioRxiv; 2025 [cited 2025 June 13]. p. 2025.03.17.643822. Available from: https://www.biorxiv.org/content/10.1101/2025.03.17.643822v2

23. Watson S, Caster O, Rochon PA, den Ruijter H. Reported adverse drug reactions in women and men: Aggregated evidence from globally collected individual case reports during half a century. EClinicalMedicine. 2019 Dec;17:100188.