

A high-throughput outcome-wide framework for profiling GLP-1RA and SGLT2i effects in type 2 diabetes

Siqi Chen^{1,2}, Bingyu Zhang^{1,2}, Huilin Tang^{2,3}, Wenjie Hu^{2,3}, Jie Hu^{2,3}, Jiajie Chen^{2,3}, Yong Chen¹⁻⁶

1 The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA

2 The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA

3 Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

4 Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA

5 Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA

6 Leonard Davis Institute of Health Economics, Philadelphia, PA, USA

Background

Type 2 diabetes (T2D) affects approximately 540 million adults worldwide, creating a clinical imperative beyond glycemic control. Randomized trials have established cardiorenal benefits of GLP-1 receptor agonists (GLP-1RAs) and sodium-glucose cotransporter-2 inhibitors (SGLT2is), while recent electronic health record studies have reported associations across many additional clinical outcomes. Existing outcome-wide studies generally analyze one treatment contrast and one outcome at a time, limiting coherent estimation of outcome-wide profiles for different drug classes and ignoring dependence shared across outcomes.

We developed a joint outcome-wide treatment-effect framework that estimates the effects of GLP-1RA and SGLT2i relative to a prespecified common reference across many outcomes. DPP-4 inhibitors (DPP-4is) were chosen as a clinically relevant active glucose-lowering comparator. Using one reference permits simultaneous estimation of GLP-1RA-versus-DPP-4i and SGLT2i-versus-DPP-4i outcome-wide profiles while sharing information across treatment groups and outcomes. These contrasts are not a direct GLP-1RA-versus-SGLT2i causal comparison and do not establish superiority of either class; a dedicated head-to-head active-comparator study would address that distinct question.

Methods

We assembled an active-comparator cohort from University of Pennsylvania Health System electronic health records. Index dates ranged from January 1, 2019, through September 30, 2024; diagnosis records from 2018–2024 were used. The index date was initiation of GLP-1RA (n=9,194), SGLT2i (n=11,657), or DPP-4i (n=8,038; reference). Eligible adults had T2D, at least one baseline-period visit, no type 1 diabetes, and no baseline end-stage renal disease or dialysis. Baseline medication history was included in the covariate adjustment. Exposure was retained from index without censoring for discontinuation or switching, consistent with an intention-to-treat-like analysis. Post-index diagnoses were observed through available 2024 data; no competing-risk model was applied.

ICD-10-CM diagnoses were mapped to category-level outcomes using HCUP CCSR v2025.1 [4]. The default CCSR category for each code was used, so multi-category assignments were not retained. All available nonconstant categories were analyzed, yielding 527 binary post-index diagnosis outcomes. The earliest diagnosis after index defined each endpoint; a 365-day pre-index CCSR indicator represented prior history, but patients with prior diagnoses were not excluded, so the endpoints are not strictly incident. All 749 baseline covariates entered the model: demographics, 11 laboratory measures, healthcare utilization, prior medications, 176 medical-history indicators, and 520 prior CCSR indicators. Laboratory missingness was median-imputed; missing visit counts and binary indicators were set to zero. Covariates were generated before index and were not reduced dimensionally. Propensity scores and standardized balance diagnostics were not used; measured confounding was addressed through covariate adjustment in the penalized model. Because the 527 outcomes may be jointly affected by unmeasured patient characteristics that are also associated with treatment selection, analyzing each outcome independently may leave shared confounding structure unaddressed. The latent-factor component is designed to represent such shared unmeasured confounding through a low-rank factor structure, while also capturing other sources of cross-outcome dependence. Incorporating these estimated latent factors into the joint model allows the framework to adjust for latent patient heterogeneity shared across outcomes and to share information during treatment-effect estimation. Mirror statistics then control the false discovery rate (FDR) at 10% across the 1,054 treatment-outcome tests.

Results

At an FDR threshold of 10%, the mirror-statistic procedure selected 125 of 1,054 treatment-outcome pairs, representing 111 unique outcomes. The selected profiles were asymmetric: 110 pairs involved SGLT2i versus DPP-4i and 15 involved GLP-1RA versus DPP-4i. All selected SGLT2i estimates were negative, whereas the GLP-1RA profile included 6 negative and 9 positive estimates.

The SGLT2i-versus-DPP-4i profile extended across cardiovascular, renal, respiratory, neuropsychiatric, musculoskeletal, metabolic, and infectious outcomes. Representative estimates included fluid and electrolyte disorders (risk difference [RD], -0.067 ; 95% CI, -0.081 to -0.052), acute renal failure (-0.039 ; -0.051 to -0.027), cardiac dysrhythmias (-0.038 ; -0.050 to -0.025), and bacterial infections (-0.037 ; -0.050 to -0.025).

The GLP-1RA-versus-DPP-4i profile was more concentrated and bidirectional. Negative estimates included fluid and electrolyte disorders (RD, -0.041 ; 95% CI, -0.054 to -0.028) and aplastic anemia (-0.015 ; -0.028 to -0.002); positive estimates included obesity (0.096 ; 0.082 to 0.110), other specified joint disorders (0.023 ; 0.013 to 0.033), and sleep-wake disorders (0.021 ; 0.007 to 0.034). This contrast does not imply that GLP-1RA is inferior to SGLT2i.

Figure 1 displays both complete DPP-4i-referenced outcome profiles, and Figure 2 summarizes estimates and 95% confidence intervals for the 125 mirror-selected pairs; neither estimates a direct GLP-1RA-versus-SGLT2i effect.

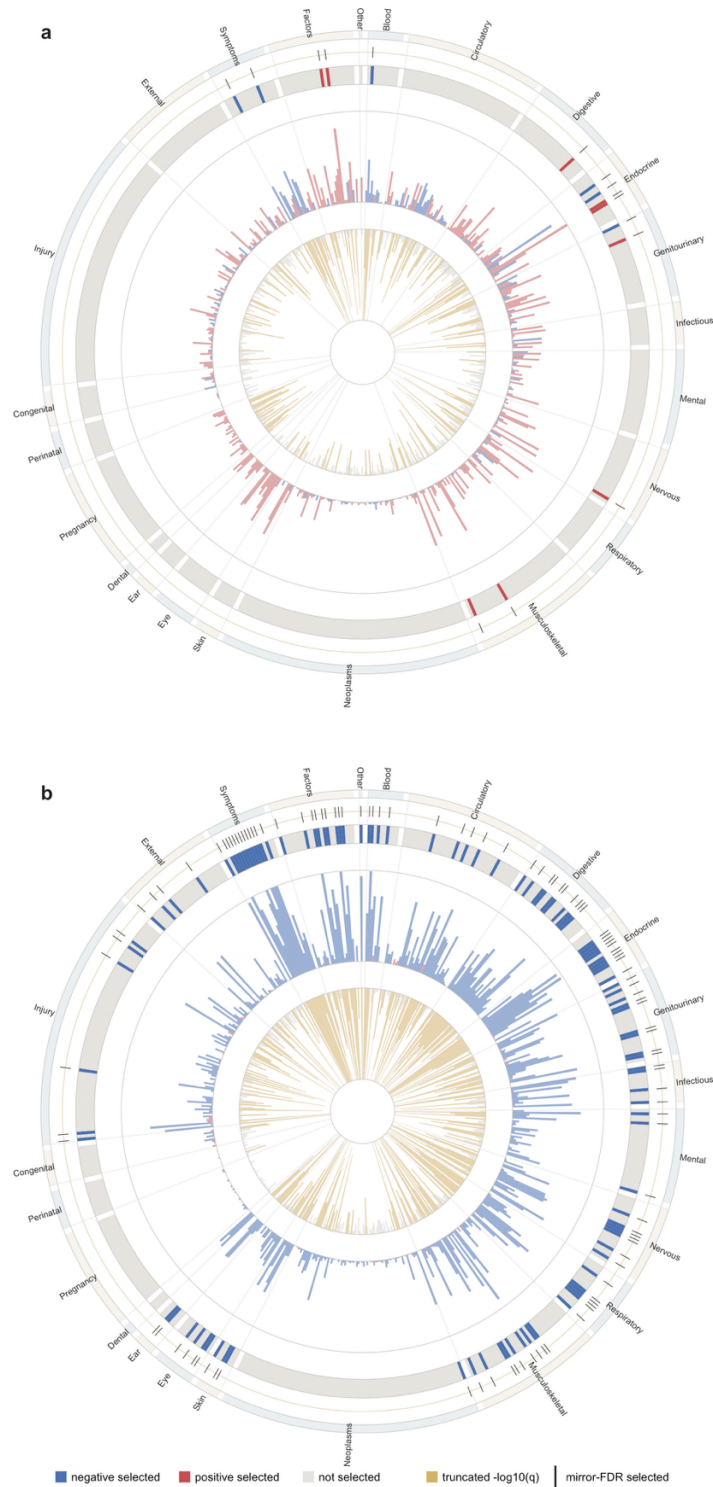
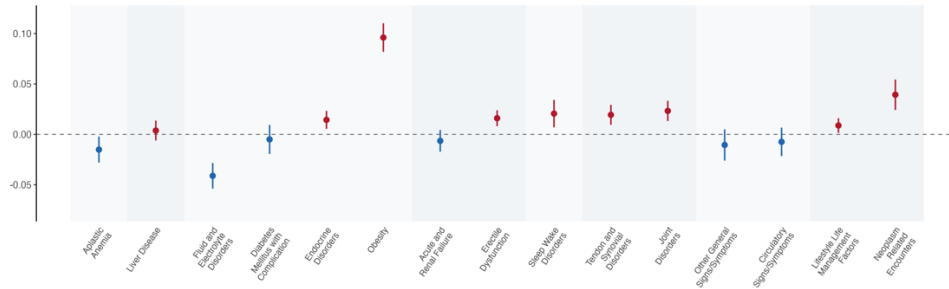


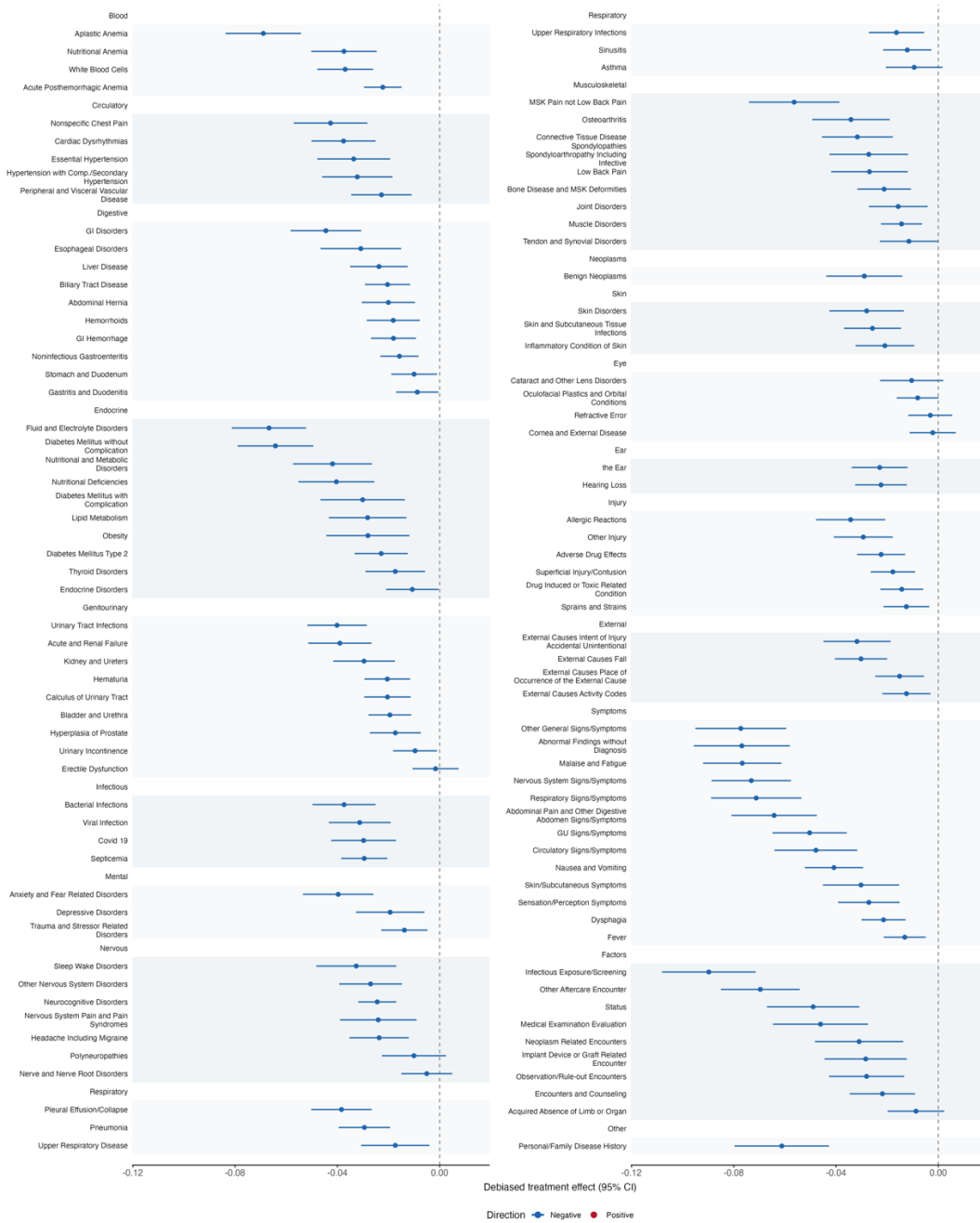
Figure 1. Circular mirror-FDR map of treatment-outcome associations across all 527 CCSR domains. Panel a: GLP-1RA vs. DPP-4i. Panel b: SGLT2i vs. DPP-4i. Outer ring indicates mirror-FDR selection status; inner bars show debiased effect magnitude; color encodes direction (blue = risk-reducing; red = risk-increasing).

a.

Debiased treatment effect (95% CI)



b.



Debiased treatment effect (95% CI)

Direction ● Negative ● Positive

Figure 2. Forest plot of debiased treatment effect estimates (95% CI) for all 125 mirror-FDR selected treatment-outcome pairs. Blue: SGLT2i vs. DPP-4i. Red: GLP-1RA vs. DPP-4i. Outcomes grouped by CCSR disease domain.

Conclusion

The proposed framework jointly estimates treatment-effect profiles across high-dimensional clinical outcomes. It combines adjustment for 749 measured covariates with a latent-factor structure for shared unmeasured confounding and cross-outcome dependence, debiased effect estimation, and mirror-statistic FDR control. Negative controls were not used in this analysis; shared unmeasured confounding across outcomes was modeled through the latent-factor component. In 28,889 adults with T2D, SGLT2i initiation was associated with a broad profile of lower risks across cardiovascular, renal, respiratory, neuropsychiatric, musculoskeletal, metabolic, and infectious outcomes relative to DPP-4i, whereas the GLP-1RA profile was more concentrated and bidirectional. Because the number and nature of unmeasured confounders cannot be fully determined, residual confounding may remain; the single-center setting may also limit generalizability. Future work should assess robustness through sensitivity analyses and validate the findings across multiple healthcare systems.

References

1. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016;375:311-22.
2. 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;373:2117-28.
3. Xie Y, Choi T, Al-Aly Z. Mapping the effectiveness and risks of GLP-1 receptor agonists. *Nat Med*. 2025;31:951-62.
4. Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. <https://hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp>