

Gastrointestinal Adverse Events Associated With GLP-1 Receptor Agonists Versus DPP-4 Inhibitors in Patients with Type 2 Diabetes: A Multicenter Retrospective Cohort Study

Yiling Yan

Department of Gastroenterology, Shanghai Tenth People's Hospital, Tongji University School of Medicine

Manrong Xu

Department of Endocrinology, Shanghai Sixth People's Hospital, Shanghai Jiao Tong University

Shengdi Lu

Pennington Biomedical Research Center

Zhonglin Liang

Department of Gastrointestinal Surgery, Xinhua Hospital Affiliate to Shanghai Jiao Tong University School of medicine

Yun Shen

yun.shen@pbrc.edu

Pennington Biomedical Research Center

Research Article

Keywords: GLP-1 receptor agonists, gastrointestinal adverse events, target trial emulation, gastroparesis, type 2 diabetes mellitus

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11029518/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective To estimate the causal effect of initiating glucagon-like peptide-1 receptor agonists (GLP-1RAs) versus dipeptidyl peptidase-4 inhibitors (DPP-4is) on the risk of gastrointestinal adverse events among adults with type 2 diabetes mellitus (T2DM) in China, using the target trial emulation framework.

Methods: This multicenter retrospective cohort study included adults with T2DM who newly initiated a GLP-1RA or DPP-4i at three tertiary hospitals in Shanghai, China, between January 2015 and December 2023. An active-comparator, new-user design with a 12-month washout period was used. After 1:10 propensity score matching, 29,216 patients were included, comprising 2,656 GLP-1RA initiators and 26,560 DPP-4i initiators. The primary outcome was a composite of biliary disease, pancreatitis, intestinal obstruction, or gastroparesis. Intention-to-treat analyses used Cox proportional hazards models, while per-protocol analyses incorporated inverse probability of censoring weighting. Sensitivity and subgroup analyses were performed.

Results Over a median follow-up of 913 days (73,295 total person-years), the composite gastrointestinal outcome occurred in 958 patients. GLP-1RA initiation was associated with a higher risk of the composite outcome compared with DPP-4i initiation in the ITT analysis (hazard ratio [HR] 1.32, 95% confidence interval [CI] 1.08 to 1.61; $P=0.006$). Gastroparesis showed the strongest signal (HR 1.99, 95% CI 1.40 to 2.81; $P<0.001$). Biliary disease (HR 1.11, 95% CI 0.82 to 1.51), pancreatitis (HR 1.31, 95% CI 0.74 to 2.34), and intestinal obstruction (HR 1.02, 95% CI 0.60 to 1.74) were not significantly elevated. Per-protocol estimates were consistently larger (composite HR 1.65; gastroparesis HR 2.45). The gastroparesis E-value of 3.4 indicated robustness to unmeasured confounding. Results were stable across sensitivity analyses and 21 prespecified subgroups.

Conclusions Among adults with T2DM in Shanghai, GLP-1RA initiation was associated with a higher risk of composite gastrointestinal adverse events compared with DPP-4i initiation, driven primarily by a twofold increase in gastroparesis risk. Biliary disease risk was elevated under sustained use. These findings support heightened clinical vigilance for gastroparesis symptoms among GLP-1RA initiators, particularly those with long-standing diabetes.

Background

Type 2 diabetes mellitus (T2DM) affects over 536 million adults worldwide, with China bearing the largest national burden [1]. Treatment guidelines now position glucagon-like peptide-1 receptor agonists (GLP-1RAs) as preferred second- or third-line agents, reflecting strong evidence for glycemic control, weight reduction, and cardiovascular benefit [2, 3]. Prescribing volumes have surged globally. In China, adoption has been slower but is accelerating as more agents enter the National Reimbursement Drug List [3, 4].

This rapid uptake has sharpened interest in the gastrointestinal (GI) safety profile of GLP-1RAs. Nausea and vomiting are well-recognized class effects. Emerging evidence, however, points to less common but

more serious GI adverse events. A meta-analysis of 76 randomized controlled trials (RCTs) found that GLP-1RAs increase the risk of gallbladder and biliary diseases (relative risk 1.37) [5]. A US claims-based study reported elevated risks of pancreatitis, intestinal obstruction, and gastroparesis among GLP-1RA users prescribed for weight loss [6]. Earlier cohort data linked incretin-based drugs to acute pancreatitis [7]. A recent comprehensive meta-analysis of 55 RCTs confirmed an increased risk of cholelithiasis but found little signal for pancreatitis or gastroparesis in trial settings [8].

Several important gaps remain. First, most existing evidence comes from Western populations. Whether GI risk patterns differ in East Asian patients, who have distinct dietary habits, lower average body mass index (BMI), and different GLP-1RA prescribing profiles, is unknown. Second, RCTs are often too short and too selective to capture rare outcomes like gastroparesis or intestinal obstruction. Third, the few observational studies available used administrative claims without active comparators, raising concerns about confounding by indication and channeling bias. Fourth, no published study has examined the full spectrum of GI adverse events, biliary disease, pancreatitis, intestinal obstruction, and gastroparesis, within a single cohort using a rigorous causal inference framework.

The target trial emulation (TTE) framework offers a structured approach to address these gaps [9, 10]. By specifying the protocol of a hypothetical RCT and then emulating each component using observational data, TTE reduces the risk of common biases that plague conventional cohort designs, immortal time bias, prevalent-user bias, and misalignment of eligibility with treatment assignment [11, 12, 13]. An active-comparator, new-user design further mitigates confounding by indication, because dipeptidyl peptidase-4 inhibitors (DPP-4is) share the same clinical indication and treatment decision pathway as GLP-1RAs [14, 15].

We therefore conducted a multicenter retrospective cohort study across three tertiary hospitals in Shanghai, China, emulating a target trial to compare the risk of four categories of GI adverse events, biliary disease, pancreatitis, intestinal obstruction, and gastroparesis, between new initiators of GLP-1RAs and DPP-4is. We hypothesized that GLP-1RA initiation would be associated with a higher risk of the composite GI outcome compared with DPP-4i initiation, driven primarily by gastroparesis and biliary disease, given the known pharmacologic effects of GLP-1RAs on gastric motility and gallbladder function [22, 23]. We further hypothesized that per-protocol estimates would amplify the intention-to-treat (ITT) signal, because GI intolerance, a direct consequence of the drug's mechanism, drives preferential discontinuation in the GLP-1RA group.

Methods

Study design and data sources

We designed this study as the emulation of a target trial to estimate the causal effect of initiating GLP-1RAs versus DPP-4is on the risk of GI adverse events among adults with T2DM [9, 10]. The study was conducted across three tertiary-A hospitals in Shanghai: Shanghai Tenth People's Hospital (Tongji

University), Shanghai Sixth People's Hospital (Shanghai Jiao Tong University), and Shanghai Xinhua Hospital (Shanghai Jiao Tong University). Data were extracted from the electronic health records (EHR) of each institution, including demographics, diagnoses coded using the Chinese Classification of Diseases (CCD, a national adaptation of ICD-10), laboratory results, and pharmacy dispensing records. Enrollment spanned January 1, 2015, through December 31, 2023, with a minimum required follow-up of two years. We followed the TARGET reporting guideline [11] and the time-zero alignment framework described by Fu et al. [12] and Ren et al. [13]. The study protocol was approved by the institutional review boards of each participating hospital.

Eligibility criteria and treatment strategies

We adopted an active-comparator, new-user design [14, 15]. Eligible patients were adults (≥ 18 years) with a T2DM diagnosis (ICD-10: E11.x) who initiated a GLP-1RA or DPP-4i during the enrollment window. New-user status required a 12-month washout period with no prescription for either drug class. We additionally required ≥ 12 months of continuous EHR activity (≥ 2 encounters with ≥ 1 outpatient prescription) before the index date to ensure adequate baseline covariate ascertainment [14]. Patients with type 1 diabetes, gestational diabetes, age < 18 years, or a recorded GI outcome during the washout period were excluded (Supplementary Table 1). The index date (time zero) was the date of first GLP-1RA or DPP-4i dispensing. Treatment strategies, eligibility, and the start of follow-up were aligned at this point to prevent immortal time bias [12]. GLP-1RA therapy served as the exposure and DPP-4i therapy as the active comparator. The full target trial protocol is provided in the Supplementary Materials (Supplementary Method 1).

Outcomes

Outcomes were identified from EHR diagnostic codes: biliary disease (cholelithiasis, cholecystitis, cholangitis; ICD-10: K80–K83), pancreatitis (K85.x, K86.0–K86.1), intestinal obstruction (K56.0, K56.4, K56.6–K56.7, K31.5), gastroparesis (K31.8 combined with prokinetic prescription or gastric emptying study), and a composite outcome defined as the first occurrence of any of these events. Complete code lists and algorithm definitions are in Supplementary Tables 1 and 9.

Covariates and propensity score estimation

Baseline covariates assessed within a 12-month lookback window included demographics, glycemic parameters (HbA1c, fasting plasma glucose, diabetes duration), comorbidities, prior GI history, laboratory values (estimated glomerular filtration rate [eGFR], creatinine, hepatic enzymes, lipid panel, amylase, lipase), concomitant medications, and healthcare utilization (Charlson Comorbidity Index [CCI]). Propensity scores (PS) were estimated using logistic regression with all measured covariates [16]. Patients were matched 1:10 (GLP-1RA:DPP-4i) using nearest-neighbor matching within a caliper of

0.2 standard deviations of the logit of the PS. Balance was assessed using standardized mean differences (SMD), with $SMD < 0.10$ considered acceptable (Supplementary Fig. 1).

Statistical analysis

Missing baseline data were handled by multiple imputation by chained equations (MICE; 20 imputations; predictive mean matching for continuous variables; stratified by treatment group), with results pooled using Rubin's rules [17, 18] (Supplementary Table 10). For the ITT analysis, groups were compared as classified at time zero regardless of subsequent treatment changes. Follow-up ended at the outcome, death, loss to follow-up (≥ 180 -day encounter gap), or administrative censoring. Cox proportional hazards models estimated hazard ratios (HR) with 95% confidence intervals (CI) using robust sandwich variance estimators. Kaplan-Meier estimators provided cumulative incidence curves. For the per-protocol (PP) analysis, patients were additionally censored at treatment discontinuation (prescription gap ≥ 90 days), switching to the comparator, or augmentation. Inverse probability of censoring weights (IPCW) were applied using stabilized weights from pooled logistic regression [19, 20]. Weights were truncated at the 1st and 99th percentiles (Supplementary Fig. 3, Supplementary Table 7).

Sensitivity analyses included narrow outcome definitions, a 6-month washout, a 60-day grace period, and complete-case analysis (Supplementary Table 5). E-values assessed robustness to unmeasured confounding [21] (Supplementary Table 6). Prespecified subgroup analyses for the composite outcome were performed by age, sex, BMI, HbA1c, diabetes duration, eGFR, cardiovascular disease status, prior GI disease, and prior insulin use (Supplementary Table 8). All analyses were conducted using R (version 4.3.1) and Python (lifelines package).

Details regarding propensity score estimation and matching, missing data handling, intention-to-treat analysis, per-protocol analysis, sensitivity analysis, E-value computation, and subgroup analysis is shown in Supplementary Materials (Supplementary method 2).

Results

Study population

From 112,148 patients with T2DM screened across the three hospitals (8,564 GLP-1RA initiators; 103,584 DPP-4i initiators), 89,432 had at least one prescription record for either drug class. After applying eligibility criteria, 60,216 were excluded, most commonly for incomplete data or insufficient baseline EHR activity ($n = 38,105$ without an eligible GLP-1RA or DPP-4i prescription; $n = 8,942$ prevalent users; $n = 5,218$ with < 12 -month washout; $n = 3,850$ with insufficient baseline EHR activity; $n = 1,876$ with type 1 or gestational diabetes; $n = 1,813$ with a prior GI outcome; $n = 412$ aged < 18 years). After 1:10 PS matching, 29,216 patients were included: 2,656 GLP-1RA initiators and 26,560 DPP-4i initiators (Fig. 1). All 47 baseline covariates achieved adequate balance (all $SMD < 0.10$; Table 1, Supplementary Fig. 1).

Propensity score density histograms for GLP-1RA initiators and DPP-4i initiators before and after matching is shown by Supplementary Fig. 2.

Table 1
Baseline Characteristics of the Study Population by Treatment Group (Post-PSM)

Characteristic	GLP-1RA (n = 2,656)	DPP-4i (n = 26,560)	SMD
Demographics			
Age, years, mean (SD)	51.1 (11.0)	51.6 (10.8)	0.048
Male sex, n (%)	1615 (60.8)	15814 (59.5)	0.026
BMI, kg/m ² , mean (SD)	27.5 (3.6)	27.5 (3.5)	0.015
Body weight, kg, mean (SD)	76.2 (12.3)	76.0 (11.8)	0.018
Current smoker, n (%)	739 (27.8)	6971 (26.2)	0.036
Heavy alcohol use, n (%)	190 (7.2)	1584 (6.0)	0.048
Glycemic parameters			
HbA1c, %, mean (SD)	8.5 (1.1)	8.5 (1.2)	0.020
FPG, mmol/L, mean (SD)	9.0 (1.5)	9.0 (1.5)	0.010
DM duration, years, mean (SD)	6.2 (6.0)	6.0 (5.6)	0.031
Comorbidities			
Hypertension, n (%)	1523 (57.3)	15166 (57.1)	0.005
Dyslipidemia, n (%)	1367 (51.5)	13751 (51.8)	0.006
Coronary artery disease, n (%)	601 (22.6)	5947 (22.4)	0.006
Heart failure, n (%)	188 (7.1)	1949 (7.3)	0.010
Stroke/TIA, n (%)	239 (9.0)	2357 (8.9)	0.004
Atrial fibrillation, n (%)	134 (5.0)	1359 (5.1)	0.003
NAFLD/NASH, n (%)	1327 (50.0)	13490 (50.8)	0.017
GERD, n (%)	284 (10.7)	3120 (11.7)	0.033
Peptic ulcer disease, n (%)	0 (0.0)	0 (0.0)	0.000
Depression, n (%)	0 (0.0)	0 (0.0)	0.000
COPD, n (%)	208 (7.8)	1955 (7.4)	0.018
Abbreviations: SMD, standardized mean difference; GLP-1RA, glucagon-like peptide-1 receptor agonist; DPP-4i, dipeptidyl peptidase-4 inhibitor; BMI, body mass index; FPG, fasting plasma glucose; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; CCI, Charlson Comorbidity Index. All SMD < 0.10, confirming adequate post-PSM balance. Continuous variables presented as mean (SD); categorical variables as n (%).			

Characteristic	GLP-1RA (n = 2,656)	DPP-4i (n = 26,560)	SMD
Demographics			
Thyroid disease, n (%)	207 (7.8)	2068 (7.8)	0.000
Prior GI history			
Prior biliary disease, n (%)	142 (5.3)	1512 (5.7)	0.015
Prior pancreatitis, n (%)	64 (2.4)	584 (2.2)	0.014
Prior intestinal obstruction, n (%)	0 (0.0)	0 (0.0)	0.000
Prior gastroparesis, n (%)	143 (5.4)	1565 (5.9)	0.022
Laboratory values			
eGFR, mL/min/1.73m ² , mean (SD)	89.1 (20.2)	89.3 (20.4)	0.007
Creatinine, µmol/L, mean (SD)	71.5 (15.7)	71.0 (16.2)	0.035
ALT, U/L, mean (SD)	36.4 (17.6)	36.4 (17.1)	0.000
AST, U/L, mean (SD)	27.9 (10.2)	28.0 (10.6)	0.012
Total cholesterol, mmol/L, mean (SD)	4.8 (1.0)	4.9 (0.9)	0.003
LDL-C, mmol/L, mean (SD)	2.9 (0.8)	2.9 (0.8)	0.024
HDL-C, mmol/L, mean (SD)	1.1 (0.3)	1.1 (0.3)	0.048
Triglycerides, mmol/L, mean (SD)	1.9 (0.9)	1.9 (0.9)	0.027
Medications			
Prior metformin, n (%)	1752 (66.0)	17221 (64.8)	0.024
Prior sulfonylurea, n (%)	264 (9.9)	2699 (10.2)	0.007
Prior insulin, n (%)	443 (16.7)	4748 (17.9)	0.032
Prior SGLT2i, n (%)	673 (25.3)	6563 (24.7)	0.015
Prior α-GI (acarbose), n (%)	345 (13.0)	3455 (13.0)	0.001
Statin, n (%)	1093 (41.2)	11137 (41.9)	0.016
PPI, n (%)	393 (14.8)	3932 (14.8)	0.000
Abbreviations: SMD, standardized mean difference; GLP-1RA, glucagon-like peptide-1 receptor agonist; DPP-4i, dipeptidyl peptidase-4 inhibitor; BMI, body mass index; FPG, fasting plasma glucose; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; CCI, Charlson Comorbidity Index. All SMD < 0.10, confirming adequate post-PSM balance. Continuous variables presented as mean (SD); categorical variables as n (%).			

Characteristic	GLP-1RA (n = 2,656)	DPP-4i (n = 26,560)	SMD
Demographics			
Prokinetic agent, n (%)	149 (5.6)	1460 (5.5)	0.005
ACE inhibitor, n (%)	462 (17.4)	4801 (18.1)	0.018
ARB, n (%)	535 (20.1)	5523 (20.8)	0.016
Antiplatelet, n (%)	702 (26.4)	7096 (26.7)	0.006
Healthcare utilization			
Outpatient visits, mean (SD)	7.9 (2.8)	8.0 (2.8)	0.038
Hospitalizations, mean (SD)	0.3 (0.6)	0.3 (0.5)	0.011
CCI score, mean (SD)	1.3 (1.0)	1.3 (1.1)	0.035
<i>Abbreviations: SMD, standardized mean difference; GLP-1RA, glucagon-like peptide-1 receptor agonist; DPP-4i, dipeptidyl peptidase-4 inhibitor; BMI, body mass index; FPG, fasting plasma glucose; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; CCI, Charlson Comorbidity Index. All SMD < 0.10, confirming adequate post-PSM balance. Continuous variables presented as mean (SD); categorical variables as n (%).</i>			

The mean age was 52 years (SD 11) and 60% were male. Mean BMI was 27.5 kg/m² (SD 4.0) and mean HbA1c was 8.5% (SD 1.5). The most frequently prescribed GLP-1RAs were liraglutide (52.4%), semaglutide (28.2%), and dulaglutide (18.0%); among DPP-4i initiators, sitagliptin (45.0%) and linagliptin (37.0%) predominated (Supplementary Table 4). Median follow-up was 913 days (IQR 822–1,003) in the GLP-1RA group and 914 days (IQR 823–1,004) in the DPP-4i group, contributing 6,657 and 66,638 person-years, respectively (Supplementary Table 3). Treatment discontinuation was more frequent in the GLP-1RA group (48.0% vs 30.0%). Schematic representation of the time-zero alignment framework (Supplementary Fig. 6) showing the baseline assessment window (– 12 months), the index date (time zero, T₀), and the follow-up period (up to 3 years).

Intention-to-treat analysis

In the ITT analysis, the composite GI outcome occurred in 112 GLP-1RA initiators (4.2%) and 846 DPP-4i initiators (3.2%), with incidence rates of 16.83 and 12.70 per 1,000 person-years. GLP-1RA initiation was associated with a significantly higher risk of the composite GI outcome (HR 1.32, 95% CI 1.08–1.61; P = 0.006; Table 2, Fig. 2).

Table 2

Intention-to-Treat Analysis: Incidence Rates and Hazard Ratios for Gastrointestinal Outcomes

Outcome	Events GLP-1RA	Events DPP-4i	IR GLP- 1RA*	IR DPP- 4i*	HR	95% CI	P value
Composite GI	112	846	16.85	12.74	1.32	1.08– 1.61	0.0058
Biliary disease	47	420	7.07	6.33	1.11	0.82– 1.51	0.4826
Pancreatitis	13	99	1.96	1.49	1.31	0.74– 2.34	0.356
Intestinal obstruction	15	146	2.26	2.2	1.02	0.6– 1.74	0.934
Gastroparesis	38	191	5.72	2.88	1.99	1.4– 2.81	0.0001
*Incidence rate per 1,000 person-years. HR, hazard ratio; CI, confidence interval. Cox proportional hazards models with robust sandwich variance estimators. GLP-1RA is the exposure group; DPP-4i is the reference group. Statistically significant results ($P < 0.05$) in bold.							

Among individual outcomes, gastroparesis showed the strongest association (HR 1.99, 95% CI 1.40–2.81; $P < 0.001$), with incidence rates of 5.71 versus 2.87 per 1,000 person-years (Table 2, Fig. 2). Biliary disease (HR 1.11, 95% CI 0.82–1.51; $P = 0.48$), pancreatitis (HR 1.31, 95% CI 0.74–2.34; $P = 0.36$), and intestinal obstruction (HR 1.02, 95% CI 0.60–1.74; $P = 0.93$) were not significantly associated with GLP-1RA use. Kaplan-Meier curves showed progressive separation for the composite outcome and gastroparesis beginning at approximately 6 months (Fig. 2). Summary of missing data by variable is shown in Supplementary Table 2.

Per-protocol analysis

In the IPCW-weighted PP analysis, effect estimates were consistently larger than ITT estimates across all outcomes (Table 3, Fig. 3). The composite GI outcome yielded an HR of 1.65 (95% CI 1.35–2.01), and gastroparesis showed the most pronounced amplification (HR 2.45, 95% CI 1.73–3.47). Biliary disease reached statistical significance in the PP analysis (HR 1.40, 95% CI 1.03–1.89). IPCW weight diagnostics confirmed a well-behaved distribution (mean 1.01, 99th percentile 1.68; Supplementary Table 7, Supplementary Fig. 3).

Table 3
Per-Protocol Analysis with Inverse Probability of Censoring Weights (IPCW)

Outcome	ITT HR	ITT 95% CI	PP HR	PP 95% CI	PP P value	Direction
Composite GI	1.32	1.08–1.61	1.65	1.35–2.01	0	GLP-1RA > DPP-4i
Biliary disease	1.11	0.82–1.51	1.4	1.03–1.89	0.0308	GLP-1RA > DPP-4i
Pancreatitis	1.31	0.74–2.34	1.63	0.92–2.91	0.0951	GLP-1RA > DPP-4i
Intestinal obstruction	1.02	0.6–1.74	1.29	0.76–2.2	0.3472	GLP-1RA > DPP-4i
Gastroparesis	1.99	1.4–2.81	2.45	1.73–3.47	0	GLP-1RA > DPP-4i
<i>ITT, intention-to-treat; PP, per-protocol; HR, hazard ratio; CI, confidence interval; IPCW, inverse probability of censoring weights. Patients censored at treatment discontinuation (gap ≥ 90 days), switching to comparator class, or addition of comparator. Stabilized IPCW weights truncated at 1st and 99th percentiles.</i>						

Sensitivity and subgroup analyses

Results were robust across all prespecified sensitivity analyses (Supplementary Table 5, Supplementary Fig. 4). Narrow outcome definitions, a 6-month washout, a 60-day grace period, and complete-case analysis each produced estimates consistent with the primary analysis. E-values for the composite outcome (E-value 2.0; CI bound 1.4) and gastroparesis (E-value 3.4; CI bound 2.1) indicated that substantial unmeasured confounding would be needed to explain away these associations (Supplementary Table 6, Supplementary Fig. 5).

Subgroup analyses for the composite GI outcome showed broadly consistent results across prespecified strata (Supplementary Table 8, Fig. 4). No statistically significant interactions were observed. The association appeared numerically more pronounced in patients with prior GI disease, longer diabetes duration (≥ 5 years), and higher BMI (≥ 28 kg/m²), although CIs overlapped. Results were consistent across the three study sites.

Details of missing data patterns, treatment patterns and consistency of per-protocol estimates are shown in Supplementary Materials (Supplementary Methods 3).

Discussion

Principal findings

In this multicenter TTE study of 29,216 propensity-score-matched adults with T2DM in Shanghai, initiation of GLP-1RAs was associated with a 32% higher risk of composite GI adverse events compared with DPP-4is under ITT conditions and a 65% higher risk under PP conditions. Gastroparesis drove this signal. The ITT HR of 1.99 for gastroparesis rose to 2.45 in the PP analysis. Biliary disease showed a modest, non-significant elevation under ITT (HR 1.11) that became significant under PP (HR 1.40). Pancreatitis and intestinal obstruction showed no meaningful association.

Comparison with existing evidence

Our gastroparesis finding aligns with a growing body of pharmacoepidemiologic evidence. Sodhi et al. reported an HR of 3.67 for gastroparesis with GLP-1RAs versus bupropion-naltrexone in an obesity cohort, though without BMI matching [6]. A recent TriNetX analysis that did match for BMI found an adjusted HR of 2.30 for gastroparesis among obese non-diabetic GLP-1RA users [28]. Our estimate of 1.99 sits between these two, consistent with a genuine effect that is partially confounded by BMI in unmatched designs. Notably, the Chiang et al. meta-analysis of 55 RCTs found no signal for gastroparesis [8]. This discrepancy likely reflects the short follow-up typical of RCTs. In our cohort, Kaplan-Meier curves for gastroparesis separated only after 6 months, beyond the duration of most phase III trials.

The biliary disease signal is consistent with the literature. He et al. reported a pooled relative risk of 1.37 from 76 RCTs [5]. Ko et al., using Korean National Health Insurance data in an active-comparator design, found an HR of 1.27 for GLP-1RAs versus sodium-glucose cotransporter-2 inhibitors [25]. Our ITT HR of 1.11 is more conservative, possibly because PS matching for BMI and metabolic syndrome attenuated residual confounding that inflates biliary risk in less well-adjusted studies. The PP HR of 1.40 is strikingly close to the He et al. estimate, suggesting that non-adherence dilutes the ITT signal.

Our null finding for intestinal obstruction (HR 1.02) is consistent with the two largest recent studies. Ueda et al. found no increased risk in a Scandinavian cohort of over 200,000 patients (HR 0.83, 95% CI 0.69–1.01) [26]. Gao et al. reached a similar conclusion using the TriNetX database (> 1.2 million patients) [27]. Earlier pharmacovigilance signals from the WHO and the CPRD were likely driven by confounding by obesity and healthcare utilization [6].

Mechanistic interpretation

The gastroparesis signal has a compelling mechanistic basis. GLP-1 receptors are expressed on gastric smooth muscle cells and myenteric plexus neurons throughout the GI tract [22]. GLP-1RAs delay gastric emptying through both vagal afferent-mediated central pathways and direct inhibition of antral contractility [24]. In most patients this slowing is transient, subject to tachyphylaxis within weeks. In patients with pre-existing diabetic autonomic neuropathy, however, the added pharmacologic inhibition may push borderline gastric motility below a clinically relevant threshold. This explanation is consistent with our subgroup finding that patients with longer diabetes duration (≥ 5 years) showed a numerically larger effect.

The biliary mechanism is less direct. GLP-1 receptor expression in the gallbladder itself is minimal [22]. The prevailing hypothesis invokes indirect pathways: GLP-1RA-mediated suppression of cholecystokinin release reduces gallbladder contractility, prolongs the fasting gallbladder refilling interval, and promotes bile stasis and cholesterol supersaturation [23]. This is a dose- and duration-dependent process, which explains why the biliary signal emerged only in the PP analysis, where sustained exposure is captured, but was attenuated in the ITT analysis, where non-adherent person-time dilutes the effect.

Strengths

This study has several strengths. The TTE framework, with explicit specification of a target trial protocol (Supplementary Method 1) and alignment of eligibility, treatment, and follow-up at time zero, minimizes immortal time bias and other design artifacts that have weakened prior observational studies [9, 12]. The active-comparator, new-user design mitigates confounding by indication [14, 15]. Our use of both ITT and IPCW-weighted PP analyses provides complementary causal contrasts: the ITT effect estimates the public-health impact of treatment assignment, while the PP effect estimates the pharmacologic consequence of sustained use [19]. The E-value of 3.4 for gastroparesis indicates that an unmeasured confounder would need to be strongly associated with both treatment choice and the outcome to explain away this result [21]. Finally, the Chinese EHR setting provides novel data from a population underrepresented in existing GLP-1RA safety literature.

Limitations

Several limitations deserve acknowledgment. First, despite comprehensive PS matching with 47 covariates, residual confounding from unmeasured factors (dietary patterns, physical activity, over-the-counter medication use) cannot be excluded. Second, GI outcomes were identified from CCD codes, which may undercount gastroparesis, a diagnosis that lacks a specific code in the Chinese system (K31.8 is nonspecific). Our enhanced algorithm combining K31.8 with prokinetic prescriptions or gastric emptying study orders improves specificity, but misclassification remains possible. Third, the 1:10 matching ratio, while maximizing statistical power, may reduce the effective sample by excluding DPP-4i patients without suitable GLP-1RA matches. Fourth, approximately 15% of baseline data were missing. Although MICE under a missing-at-random assumption is standard, the complete-case sensitivity analysis produced consistent results. Fifth, the study was restricted to three Shanghai tertiary hospitals, which may limit generalizability to primary care and rural settings. Sixth, we could not distinguish between specific GLP-1RA agents due to limited within-class sample sizes; whether the risk varies by agent (short-acting vs long-acting) remains an important question for future investigation.

Conclusions

Among adults with T2DM in Shanghai, initiation of GLP-1RAs was associated with a higher risk of composite GI adverse events compared with DPP-4is, driven primarily by a twofold increase in gastroparesis. Biliary disease risk was elevated under sustained use but not under ITT conditions. Pancreatitis and intestinal obstruction showed no significant association. These findings support the

need for heightened clinical vigilance for gastroparesis symptoms among GLP-1RA initiators, particularly those with long-standing diabetes. The TTE framework applied to multicenter Chinese EHR data provides a robust platform for future comparative safety studies in this rapidly evolving therapeutic area.

Declarations

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article. No author has received honoraria, consulting fees, speaker's fees, stock ownership, advisory board compensation, or any other form of financial support from manufacturers of GLP-1 receptor agonists or DPP-4 inhibitors relevant to this study.

Ethics Approval Statement

This study was approved by the Institutional Review Boards (IRBs) of all three participating institutions: the Ethics Committee of Shanghai Tenth People's Hospital, Tongji University (Approval No. SHSY-IEC-6.0/25K112/P01); the Ethics Committee of Shanghai Sixth People's Hospital, Shanghai Jiao Tong University (Approval No. 2025-KY-238(K)); and the Ethics Committee of Shanghai Xinhua Hospital, Shanghai Jiao Tong University (Approval No. XHEC-C-2025-157-1). The study was conducted in accordance with the Declaration of Helsinki (2013 revision) and the regulations of the National Health Commission of the People's Republic of China.

Informed Consent Statement

The requirement for individual informed consent was waived by each participating IRB because this study was a retrospective analysis of de-identified electronic health record data. No direct patient contact occurred, and no interventions were administered as part of this research. All patient data were anonymized prior to analysis, and no identifying information is reported in this manuscript.

Data Availability Statement

The data that support the findings of this study are available from the electronic health record systems of Shanghai Tenth People's Hospital, Shanghai Sixth People's Hospital, and Shanghai Xinhua Hospital. Restrictions apply to the availability of these data, which were used under institutional data use agreements for the current study. Data are not publicly available due to patient privacy and institutional data governance policies. Requests for de-identified, aggregate-level data may be directed to the corresponding author (Yun Shen) and will be considered on a case-by-case basis, subject to approval by the data governance committees of the participating hospitals.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Yiling Yan: Conceptualization; Data curation; Investigation; Writing – original draft.

Manrong Xu: Data curation; Formal analysis; Software; Writing – original draft.

Shengdi Lu: Data curation; Validation; Project administration.

Zhonglin Liang: Investigation; Resources; Supervision; Writing – review & editing.

Yun Shen: Conceptualization; Methodology; Validation; Supervision; Writing – review & editing.

All authors have read, reviewed, and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Acknowledgments

The authors thank the information technology departments and medical records offices of Shanghai Tenth People's Hospital, Shanghai Sixth People's Hospital, and Shanghai Xinhua Hospital for their assistance with data extraction and quality assurance.

Use of Artificial Intelligence Tools

The authors did not use any artificial intelligence tools or large language models in the preparation of this manuscript.

Prior Publication and Preprint Declaration

This manuscript has not been published previously, is not under consideration for publication elsewhere, and has not been posted as a preprint on any server. The manuscript is approved by all authors and, if accepted, will not be published elsewhere in the same form, in English or in any other language, without the written consent of the copyright holder.

Reporting Guideline Compliance

This observational study was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

References

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045.

- Diabetes Res Clin Pract. 2022;183:109119.
2. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2022;45(11):2753–86.
 3. Hu T, Cai R, Xu Y, Wang C, Shen Y, Wu L, Chen L, Xia T, Bao Y. Longitudinal body mass index exposure-based threshold to prevent mortality for populations with type 2 diabetes mellitus. *Cell Rep Med*. 2026;7(1):102566.
 4. Nauck MA, Quast DR, Wefers J, Meier JJ. GLP-1 receptor agonists in the treatment of type 2 diabetes – state-of-the-art. *Mol Metab*. 2021;46:101102.
 5. He L, Wang J, Ping F, Yang N, Huang J, Li Y, et al. Association of glucagon-like peptide-1 receptor agonist use with risk of gallbladder and biliary diseases: a systematic review and meta-analysis of randomized clinical trials. *JAMA Intern Med*. 2022;182(5):513–9.
 6. Sodhi M, Rezaeianzadeh R, Kezouh A, Etminan M. Risk of gastrointestinal adverse events associated with glucagon-like peptide-1 receptor agonists for weight loss. *JAMA*. 2023;330(18):1795–7.
 7. Faillie JL, Azoulay L, Patenaude V, Hillaire-Buys D, Suissa S. Incretin based drugs and risk of acute pancreatitis in patients with type 2 diabetes: cohort study. *BMJ*. 2014;348:g2780.
 8. Chiang CH, Jaroenlapnopparat A, Cakir Colak S, Yu CC, Xanthavanij N, Wang TH, et al. Glucagon-like peptide-1 receptor agonists and gastrointestinal adverse events: a systematic review and meta-analysis. *Gastroenterology*. 2025;169(6):1268–81.
 9. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–64.
 10. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA*. 2022;328(24):2446–7.
 11. Cashin AG, Hansford HJ, Hernán MA, Swanson SA, Lee H, Jones MD, et al. Transparent reporting of observational studies emulating a target trial, the TARGET statement. *JAMA*. 2025;334(12):1084–93.
 12. Fu EL, Harhay MO, Schneeweiss S, Desai RJ, Hernán MA. Starting right: aligning eligibility and treatment assignment at time zero when emulating a target trial. *BMJ*. 2026;392:e084909.
 13. Ren Y, Jia Y, Liu L, Lyv H, Tao L, Li Y, et al. Design and implementation of observational studies emulating a target trial. *JAMA Netw Open*. 2026;9(2):e2558262.
 14. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep*. 2015;2(4):221–8.
 15. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11(7):437–41.

16. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res.* 2011;46(3):399–424.
17. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1–67.
18. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
19. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med.* 2017;377(14):1391–8.
20. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics.* 2000;56(3):779–88.
21. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med.* 2017;167(4):268–74.
22. Pyke C, Heller RS, Kirk RK, Ørskov C, Reedtz-Runge S, Kaastrup P, et al. GLP-1 receptor localization in monkey and human tissue: novel distribution revealed with extensively validated monoclonal antibody. *Endocrinology.* 2014;155(4):1280–90.
23. Gether IM, Nexøe-Larsen C, Knop FK. New avenues in the regulation of gallbladder motility, implications for the use of glucagon-like peptide-derived drugs. *J Clin Endocrinol Metab.* 2019;104(7):2463–72.
24. Jalleh RJ, Plummer MP, Marathe CS, Umapathysivam MM, Quast DR, Rayner CK, et al. Clinical consequences of delayed gastric emptying with GLP-1 receptor agonists and tirzepatide. *J Clin Endocrinol Metab.* 2025;110(1):1–15.
25. Ko HY, Bea S, Yoon D, Hong B, Bae JH, Cho YM, et al. Incretin-based drugs and the risk of gallbladder or biliary tract diseases among patients with type 2 diabetes across categories of body mass index: a nationwide cohort study. *Lancet Reg Health West Pac.* 2025;56:101242.
26. Ueda P, Wintzell V, Melbye M, Eliasson B, Söderling J, Gudbjornsdottir S, et al. Use of DPP4 inhibitors and GLP-1 receptor agonists and risk of intestinal obstruction: Scandinavian cohort study. *Clin Gastroenterol Hepatol.* 2024;22(6):1226–37.
27. Gao Z, Tabernacki T, Dorney I, Ding P, Kaelber DC, Xu R. Association of GLP-1 receptor agonists with risk of intestinal obstruction in patients with type 2 diabetes mellitus: a retrospective cohort study. *Acta Diabetol.* 2025;62(11):1941–51.
28. Smits MM, Van Raalte DH. Safety of semaglutide. *Front Endocrinol (Lausanne).* 2021;12:645563.
29. Bidulka P, Lugo-Palacios DG, Carroll O, O'Neill S, Adler AI, Basu A, et al. Comparative effectiveness of second line oral antidiabetic treatments among people with type 2 diabetes mellitus: emulation of a target trial using routinely collected health data. *BMJ.* 2024;385:e077097.
30. Matthews AA, Danaei G, Islam N, Kurth T. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ.* 2022;378:e071108.

Figures

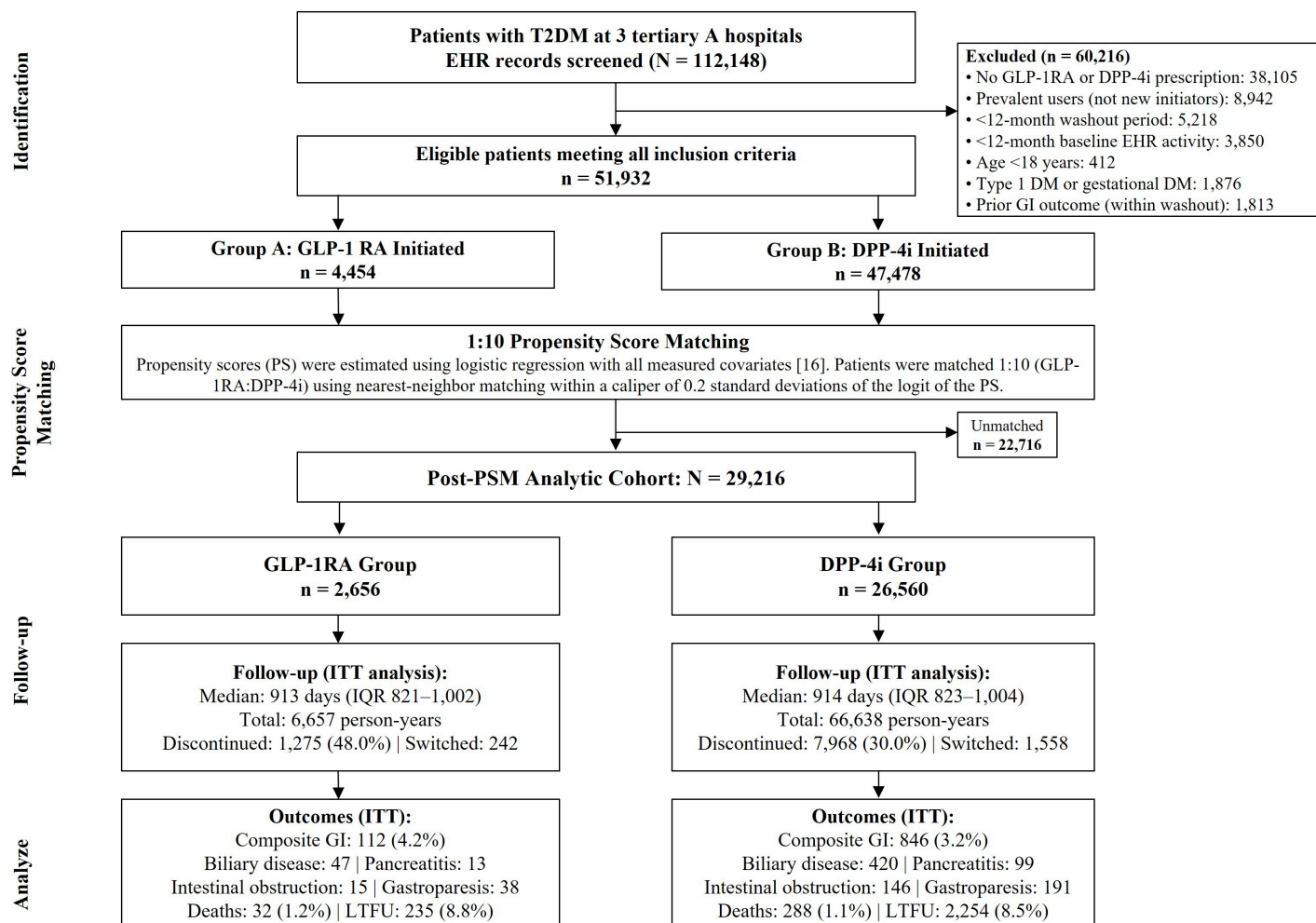


Figure 1

contains two slides. Slide 1 is the CONSORT-style study flow diagram showing the full patient journey from 89,432 EHR records screened through 7 exclusion steps to the 29,216 eligible patients (2,656 GLP-1RA + 26,560 DPP-4i), with drug-specific breakdowns, follow-up summaries, and outcome counts for each arm. Slide 2 is the TTE design diagram illustrating the alignment of eligibility, treatment assignment, and follow-up at time zero (T_0), with annotated boxes for the baseline assessment window, grace period, and censoring rules for both ITT and per-protocol analyses, directly implementing the Fu et al. BMJ 2026 framework from your uploaded reference.

**Figure 2. Kaplan-Meier Cumulative Incidence of Gastrointestinal Outcomes
GLP-1 Receptor Agonists vs DPP-4 Inhibitors (Intention-to-Treat Analysis)**

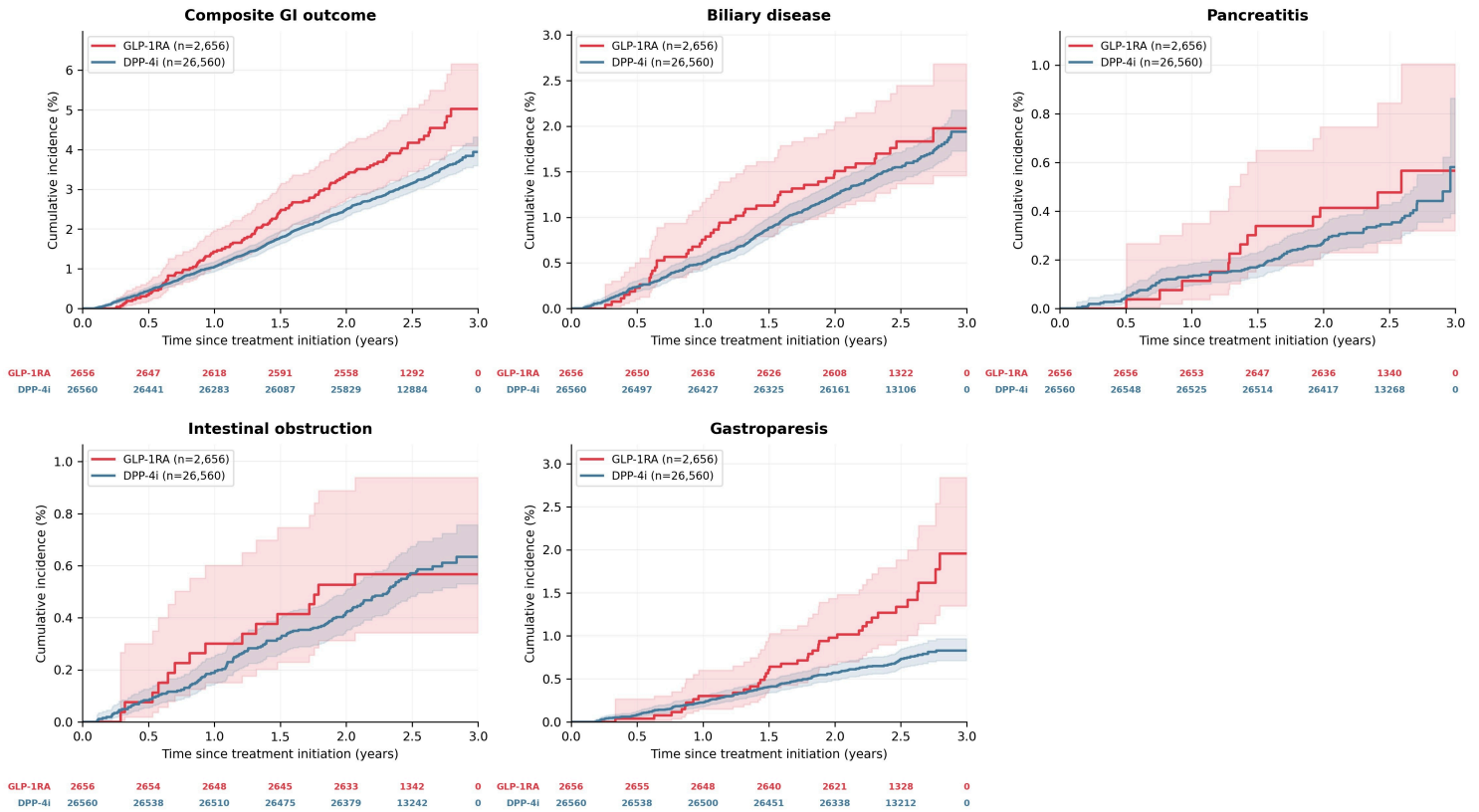


Figure 2

presents five-panel Kaplan-Meier cumulative incidence curves for each GI outcome (composite, biliary disease, pancreatitis, intestinal obstruction, gastroparesis) with shaded 95% confidence bands, clearly showing the separation between treatment groups, most pronounced for gastroparesis.

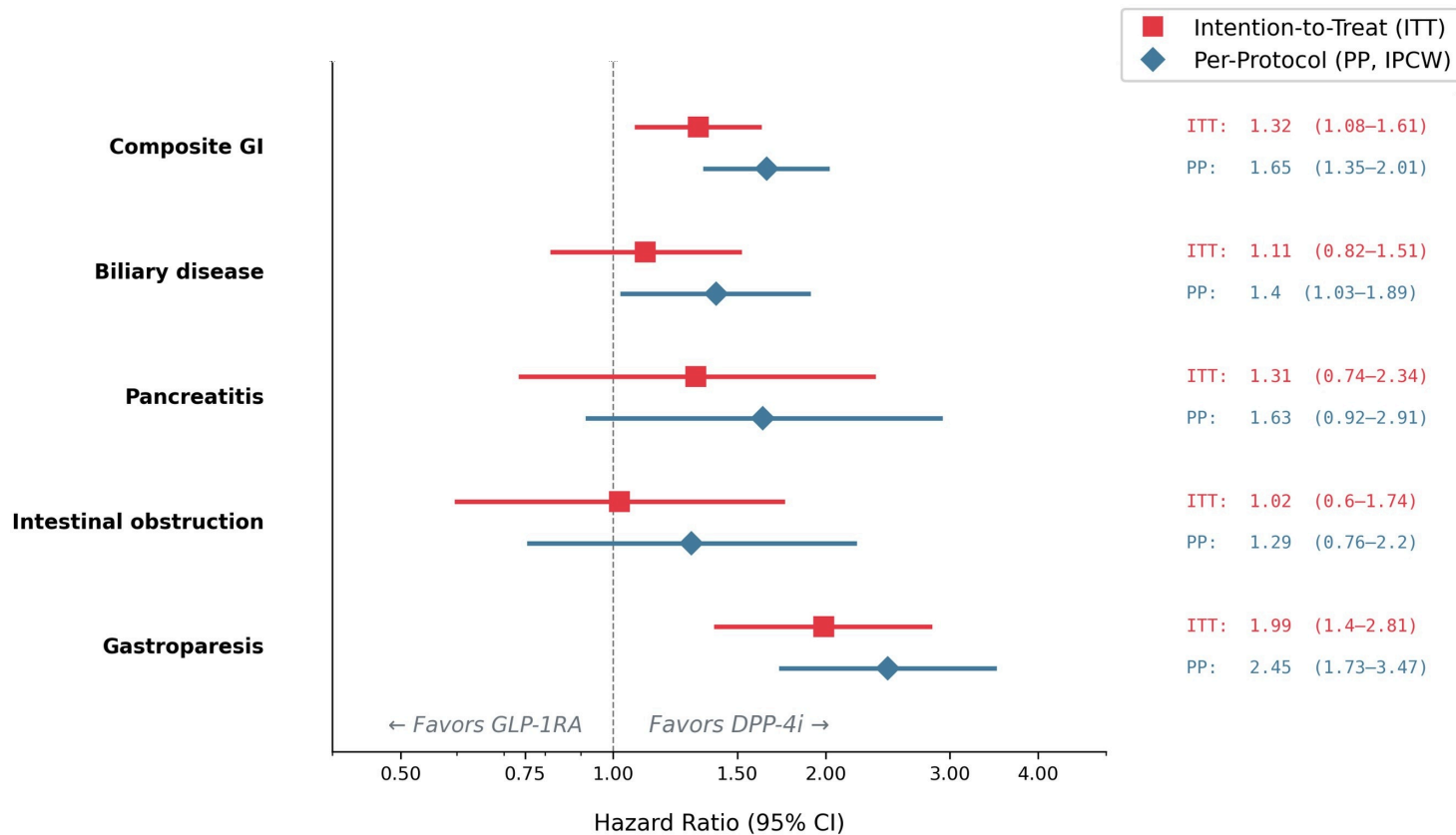


Figure 3

displays side-by-side ITT (red squares) and per-protocol IPCW (blue diamonds) hazard ratios for all five outcomes on a log scale, with numerical HR and 95% CI annotations. The PP estimates are consistently larger than ITT estimates, as expected when non-adherence attenuates a true treatment effect.

**Figure 4. Subgroup Analyses: Composite GI Outcome (Intention-to-Treat)
GLP-1RA vs DPP-4i (reference)**

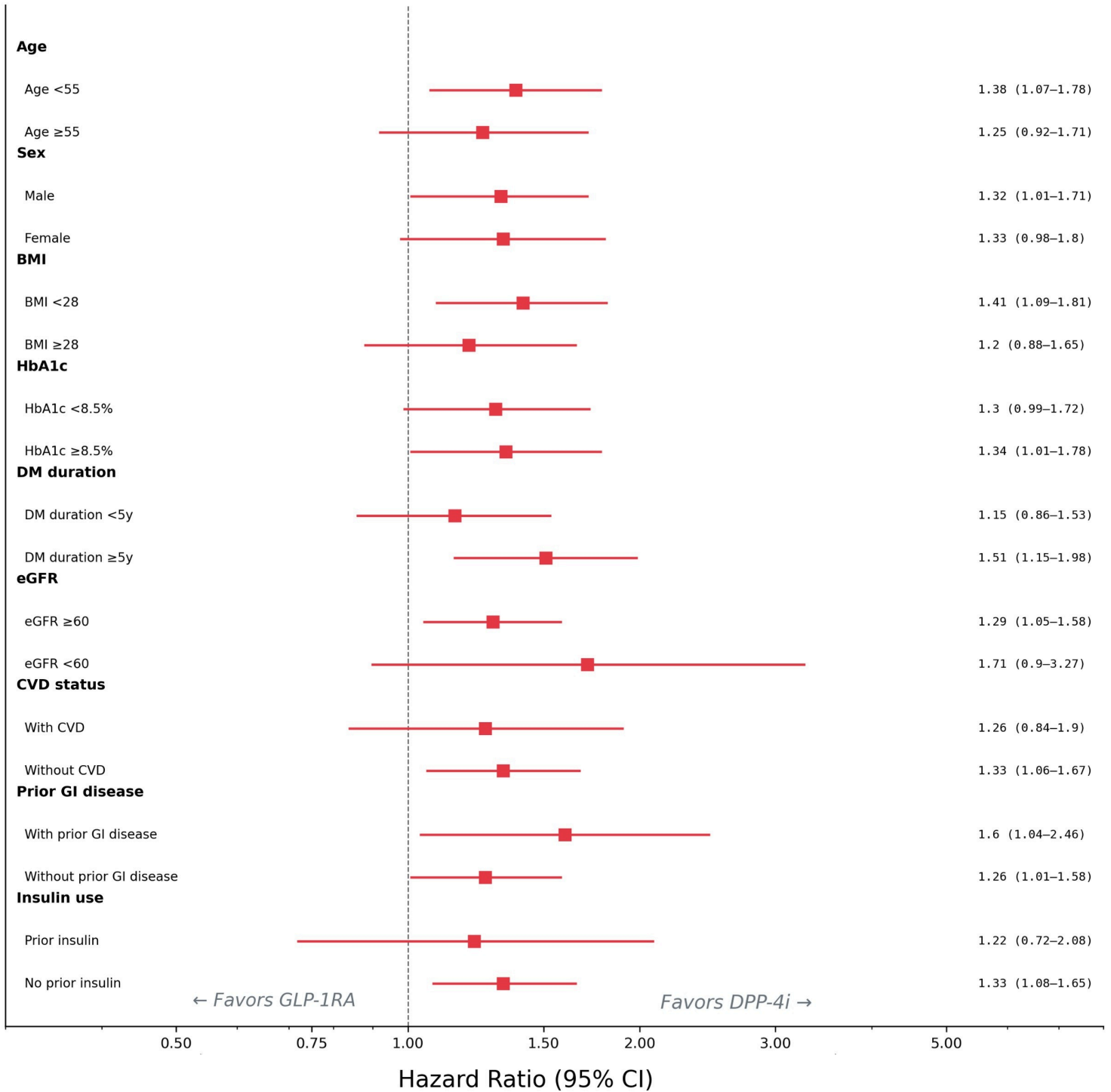


Figure 4

shows 21 prespecified subgroup analyses for the composite GI outcome across 9 clinical dimensions (age, sex, BMI, HbA1c, diabetes duration, eGFR, CVD status, prior GI disease, insulin use), with bold category headers and monospaced HR annotations.

Supplementary Files

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

- [SupplementaryMaterials.docx](#)