Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gallstone Occurrence Versus Clinically Significant Biliary Disease With GLP-1 Receptor Agonists in Type 2 Diabetes and Obesity: A Propensity-Matched Cohort Study With Phenotype and Quantitative-Bias Analysis

Chidera Onwuzo

chionwuzo@gmail.com

SUNY Upstate Medical University Hospital

Huda Jaffar

SUNY Upstate Medical University Hospital

Alvina FNU

SUNY Upstate Medical University Hospital

Hope Marshall

SUNY Upstate Medical University Hospital

Chibuike Maduka

I.Horbachevsky Ternopil National Medical University

Abdelkaader Chaar

SUNY Upstate Medical University Hospital

Mohamed Eisa

SUNY Upstate Medical University Hospital

Research Article

Keywords: cholelithiasis, GLP-1 receptor agonists, obesity, type 2 diabetes mellitus, propensity score matching, gallbladder disease

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10321029/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

Purpose

GLP-1 receptor agonists (GLP-1 RAs) are widely used for type 2 diabetes mellitus (T2DM) and obesity and have been linked to gallstone disease. Whether gallstone occurrence translates into clinically significant biliary disease, and whether this signal is driven by diabetes or obesity, remains unclear.

Methods

We performed a retrospective propensity-matched cohort study using the TriNetX network. Adults aged 18 to 65 with T2DM or obesity and recorded body mass index (BMI) were grouped by GLP-1 RA exposure and matched one to one on 20 covariates, including BMI. Six biliary outcomes were followed for five years. We reported odds ratios for cumulative occurrence, Cox hazard ratios and Kaplan-Meier survival for event timing, phenotype subgroups, and E-values for confounding.

Results

We matched 721,898 pairs. GLP-1 RA use was associated with slightly higher odds of cholelithiasis (OR 1.11) and choledocholithiasis (OR 1.14), and lower odds of cholecystitis (0.94), cholangitis (0.51), cholecystectomy (0.62), and ERCP (0.46). Hazard ratios were below 1 for all outcomes, including stones, and five-year cholelithiasis incidence was nearly identical. Higher stone odds appeared only in obesity without diabetes and were absent in T2DM-only and combined groups. Inverse associations for inflammatory and procedural outcomes appeared in every subgroup. E-values were large for inverse associations and small for stone associations.

Conclusions

Increased gallstone occurrence with GLP-1 RAs was modest, limited to obesity without diabetes, and not evident after accounting for event timing. It did not correspond to more clinically significant biliary disease. Inverse findings should be interpreted as associations, not benefit.

Introduction

Glucagon-like peptide-1 receptor agonists (GLP-1 RAs) are widely used in the management of type 2 diabetes mellitus (T2DM) and obesity. In addition to effective glycemic control, multiple studies have demonstrated benefits including reduced major cardiovascular events, improved metabolic profiles, and clinically meaningful weight reduction, leading to more frequent use of these in cardiometabolic populations [1–3].

The expanding use of GLP-1 RAs has also raised questions about their hepatobiliary safety profile. Patients with T2DM and obesity are already at increased risk for gallbladder disease because of insulin resistance, metabolic dysfunction, altered bile composition, and excess body weight. GLP-1 RA therapy may further modify this risk through several proposed mechanisms. GLP-1 signaling may reduce gallbladder contractility, alter gastrointestinal motility, and impair cholecystokinin-mediated gallbladder emptying, leading to bile stasis [4–6]. The rapid weight loss associated with GLP-1 RAs is also believed to lead to supersaturation of bile with cholesterol leading to a higher incidence of gallstones formation [6–7]. Prior studies have suggested that biliary risk may be more pronounced with higher doses and longer duration of GLP-1 RA therapy [7].

However, the clinical significance of GLP-1 RA-associated gallstone occurrence remains uncertain. A higher frequency of cholelithiasis or choledocholithiasis does not necessarily indicate progression to downstream inflammatory, infectious, or procedural outcomes such as cholecystitis, cholangitis, cholecystectomy, or endoscopic retrograde cholangiopancreatography. This distinction is important because stone formation, inflammatory complications, and biliary interventions may reflect different stages of disease and may be influenced by different biological and clinical factors.

GLP-1 RAs may also have systemic effects that could theoretically influence progression from gallstone formation to clinically significant biliary disease. Improved glycemic control, weight reduction, and emerging anti-inflammatory effects of GLP-1-based therapies may affect inflammatory or infectious pathways, although this relationship remains unclear and should not be assumed to be causal [8–9]. Whether biliary risk differs between patients treated for T2DM, obesity, or both conditions also remains incompletely understood.

Given these uncertainties, further real-world data are needed to clarify whether GLP-1 RA exposure is associated with increased gallstone occurrence, whether this translates into clinically significant biliary disease, and whether these associations differ by metabolic phenotype. In this study, we used a large global federated database to evaluate biliary outcomes among adults with T2DM or obesity exposed to GLP-1 RAs compared with matched non-users, incorporating cumulative outcome measures, time-to-event analysis, and phenotype-based subgroup analyses.

Methods

Data source

We used the TriNetX Global Collaborative Network, a federated database of de-identified electronic health records from 177 healthcare organizations [10]. The platform returns aggregate statistics and does not release patient-level records. Because the data are de-identified and aggregated, the study did not require institutional review board approval.

Study design and population

This was a retrospective, propensity-matched cohort study. We included adults aged 18 to 65 years with a diagnosis of type 2 diabetes (ICD-10 E11) or obesity (E66.9) and at least one recorded BMI. We excluded patients with prior bariatric surgery (Z98.84) or pregnancy. The exposed group received any GLP-1 RA (tirzepatide, semaglutide, dulaglutide, liraglutide, exenatide, lixisenatide, or albiglutide) on or after the qualifying diagnosis. The comparison group had the same diagnoses and no record of any GLP-1 RA.

Exposure, index date, and follow-up

The index date was the first GLP-1 RA record for the exposed group and the qualifying diagnosis for the comparison group. Outcomes were counted from one day after the index date through five years (1,825 days). For each outcome, patients who already had that outcome before follow-up were excluded, so that only new (incident) events were analyzed.

Outcomes

We assessed six biliary outcomes. Cholelithiasis (K80) and choledocholithiasis (K80.50) represented stones. Cholecystitis (K81) and cholangitis (K83.0) represented inflammation and obstruction. Cholecystectomy and ERCP represented procedures, which mark biliary disease serious enough to need intervention.

Propensity-score matching

We matched the two groups one to one by propensity score on 20 covariates, covering demographics (age, sex, race and ethnicity), cardiometabolic conditions (hypertension, heart failure, cerebrovascular disease, ischemic heart disease, chronic kidney disease, hyperlipidemia, and hepatic steatosis), substance use (nicotine, alcohol, and opioids), glycemic markers (hemoglobin A1c testing and insulin use), and BMI. BMI was included because body weight is a strong confounder of biliary disease. Balance was judged by the standardized mean difference (SMD), with values below 0.10 considered acceptable.

Statistical analysis

We used two measures that answer different questions. Odds ratios and risk ratios describe how often an outcome occurred over the five-year window (cumulative occurrence). Kaplan-Meier survival analysis, with the log-rank test and Cox hazard ratios, describes how quickly events occurred and accounts for differences in follow-up time by censoring each patient at the last recorded activity. Reporting both is deliberate. An odds ratio above 1 with a hazard ratio below 1 indicates that an apparent excess reflects more cumulative diagnoses rather than a faster rate of events.

Phenotype subanalysis

To separate the effect of diabetes from the effect of obesity, we repeated the analysis in three mutually exclusive groups using identical methods: type 2 diabetes without an obesity diagnosis (T2DM only), obesity without type 2 diabetes (obesity only), and both conditions together. Each group was matched independently.

E-value sensitivity analysis

The E-value was calculated to assess how robust the observed associations are to unmeasured confounding. It is the minimum strength of association, on the risk-ratio scale, that an unmeasured factor would need with both the exposure and the outcome, beyond the measured covariates, to fully explain away an observed association. A large E-value means the result is hard to overturn; a value near 1 means it is fragile.

E-values do not directly quantify bias from imperfect balance in a covariate that was included in the model, but they provide a useful benchmark for the size of confounding that would be needed to nullify a result. On this basis, the E-value can be interpreted as reflecting robustness to residual confounding from both unmeasured factors and incompletely balanced covariates, including BMI. Because all outcomes were rare (under 5 percent), the odds and hazard ratios approximate the risk ratio, so E-values were computed from each estimate and from the confidence limit nearest 1. Because GLP-1 RA users remained heavier after matching, any leftover confounding pushes every estimate upward, which makes the inverse associations conservative and the elevated associations fragile.

Analyses were performed on the TriNetX platform. A two-sided p value below 0.05 was considered significant. Given the number of comparisons, we focused on the size and consistency of effects rather than on individual p values.

General artificial intelligence tools, including ChatGPT by OpenAI, were used to assist with language editing, manuscript organization, and generation of tables and figures. All AI-assisted content, including tables and figures, was reviewed and verified by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Results

Cohort and matching

After matching, the pooled cohort had 721,898 GLP-1 RA users and the same number of matched non-users. Every covariate except BMI was balanced to an SMD below 0.05. BMI remained imbalanced (37.8 versus 34.5 kg/m², SMD 0.43), with GLP-1 RA users heavier (Table 1). This gap persisted in every analysis and reflects the fact that GLP-1 RA users are heavier than any available comparison group. Its

direction, treated patients being heavier, means residual confounding pushes all estimates upward, which we address in the sensitivity analysis.

Table 1
Selected baseline characteristics of the pooled cohort after propensity-score matching.

Characteristic	GLP-1 RA	No GLP-1 RA	SMD before	SMD after
Current age, years	51.1	51.3	0.250	0.023
Female, %	60.0	59.6	0.148	0.008
White, %	62.2	62.7	0.136	0.010
Black, %	21.4	20.9	0.059	0.012
Hypertension, %	25.6	25.7	0.448	0.003
Heart failure, %	2.7	2.7	0.103	< 0.001
Chronic kidney disease, %	3.1	3.1	0.098	< 0.001
Hyperlipidemia, %	11.5	11.3	0.338	0.007
HbA1c testing, %	17.4	16.6	0.539	0.021
Insulin use, %	8.8	8.9	0.266	0.007
BMI, kg/m²	**37.8**	**34.5**	**0.409**	**0.426**

BMI, body mass index; SMD, standardized mean difference. Every covariate was balanced to an SMD below 0.05 after matching except BMI.

Primary analysis

In the pooled cohort, GLP-1 RA use was associated with slightly higher odds of the stone outcomes, cholelithiasis (OR 1.11) and choledocholithiasis (OR 1.14), and with lower odds of cholecystitis (0.94), cholangitis (0.51), cholecystectomy (0.62), and ERCP (0.46) (Table 2, Fig. 1).

Time-to-event analysis changed the picture for stones. Hazard ratios were below 1 for every outcome, including cholelithiasis (HR 0.90) and choledocholithiasis (HR 0.92). The higher odds of cholelithiasis therefore reflect more cumulative diagnoses rather than a faster event rate. Once follow-up time was taken into account, the five-year cumulative incidence of cholelithiasis was almost identical in the two groups (4.40 versus 4.44 percent, Fig. 2). A raised odds ratio with a hazard ratio below 1 is the pattern expected when one group is followed and tested more closely than the other; GLP-1 RA users, who attend care regularly, accumulate more recorded diagnoses while their rate over time is no higher.

Table 2
Primary analysis: biliary outcomes in the pooled cohort (GLP-1 RA versus no GLP-1 RA).

Outcome	Events (GLP-1 / no GLP-1)	Odds ratio (95% CI)	Hazard ratio (95% CI)
Cholelithiasis	14,373 / 13,154	1.107 (1.081–1.134)	0.897 (0.876–0.919)
Choledocholithiasis	3,118 / 2,753	1.135 (1.079–1.195)	0.923 (0.877–0.972)
Cholecystitis	4,692 / 5,003	0.940 (0.903–0.978)	0.767 (0.737–0.798)
Cholangitis	468 / 912	0.513 (0.459–0.573)	0.422 (0.377–0.471)
Cholecystectomy	1,739 / 2,798	0.623 (0.586–0.661)	0.530 (0.499–0.562)
ERCP	977 / 2,142	0.455 (0.422–0.491)	0.386 (0.358–0.416)

Red indicates a significant increase; blue indicates a significant decrease. All events are incident over five years.

Survival analysis

The survival analysis agreed with the odds ratios for cholecystitis, cholangitis, cholecystectomy, and ERCP, where five-year event-free survival was higher in GLP-1 RA users, meaning fewer events (Table 3, Fig. 2). For the stone outcomes it diverged: despite odds ratios above 1, event-free survival was nearly identical between groups and hazard ratios were below 1, which indicates no faster rate of new stones once event timing was taken into account.

Table 3
Five-year Kaplan-Meier event-free survival, pooled cohort.

Outcome	GLP-1 RA	No GLP-1 RA	Hazard ratio (95% CI)	Log-rank p
Cholelithiasis	95.60%	95.56%	0.897 (0.876–0.919)	< 0.001
Choledocholithiasis	99.09%	99.10%	0.923 (0.877–0.972)	< 0.001
Cholecystitis	98.60%	98.41%	0.767 (0.737–0.798)	< 0.001
Cholangitis	99.86%	99.72%	0.422 (0.377–0.471)	< 0.001
Cholecystectomy	99.52%	99.30%	0.530 (0.499–0.562)	< 0.001
ERCP	99.70%	99.48%	0.386 (0.358–0.416)	< 0.001

Phenotype subanalysis

The phenotype subgroups located the stone signal (Table 4, Fig. 3). Higher odds of cholelithiasis and choledocholithiasis appeared only in **obesity without diabetes** (OR 1.11 and 1.14). In the T2DM-only group both stone outcomes were null to lower (0.96 and 0.94), and in patients with both conditions they

were null (0.99 and 0.99). The higher odds of stones are therefore a feature of obesity without diabetes and are absent whenever diabetes is present.

The inverse associations for cholecystitis, cholangitis, cholecystectomy, and ERCP were present in all three subgroups and were larger in the T2DM-only group (ERCP OR 0.32, cholangitis OR 0.34).

Table 4
Odds ratios (95% CI) by metabolic phenotype.

Outcome	Pooled	T2DM only	Obesity only	T2DM and obesity
Cholelithiasis	1.107 (1.081–1.134)	0.958 (0.915–1.004)	1.111 (1.061–1.164)	0.992 (0.955–1.030)
Choledocholithiasis	1.135 (1.079–1.195)	0.940 (0.843–1.048)	1.135 (1.038–1.240)	0.990 (0.912–1.075)
Cholecystitis	0.940 (0.903–0.978)	0.847 (0.781–0.919)	0.886 (0.822–0.954)	0.853 (0.800–0.908)
Cholangitis	0.513 (0.459–0.573)	0.338 (0.275–0.416)	0.443 (0.335–0.586)	0.545 (0.463–0.641)
Cholecystectomy	0.623 (0.586–0.661)	0.541 (0.478–0.613)	0.561 (0.502–0.628)	0.659 (0.598–0.725)
ERCP	0.455 (0.422–0.491)	0.315 (0.270–0.368)	0.400 (0.340–0.471)	0.522 (0.466–0.584)

Red, significant increase; blue, significant decrease; grey, not significant. Matched pairs: pooled 721,898; T2DM only 255,593; obesity only 234,202; both 230,296.

Sensitivity analysis (E-values)

E-values separated the two halves of the result (Table 5). The inverse associations had large E-values, reaching 3.8 for ERCP in the pooled cohort and 5.8 in the T2DM-only group, and 3.3 to 5.4 for cholangitis. For ERCP, an E-value of 3.8 means that an unmeasured or residual confounder would need to be associated with both GLP-1 RA use and ERCP by a risk ratio of about 3.8 each, at the same time and beyond the measured covariates, to fully explain away the association. A confounder strongly linked to only one of the two would not be enough. A residual difference of roughly three BMI points is unlikely to reach that level, and because the leftover confounding here works against the inverse direction, these estimates are conservative.

The stone associations had small E-values, about 1.5. Since residual confounding pushes these estimates upward, an already small elevation could be partly or fully explained by the leftover BMI gap, so the stone findings are reported as modest and provisional.

Table 5
E-values for the odds-ratio estimate by analysis (confidence-limit E-value in parentheses).

Outcome	Pooled	T2DM only	Obesity only	T2DM and obesity
Cholelithiasis	1.45 (1.38)	1.26 (1)	1.46 (1.32)	1.1 (1)
Choledocholithiasis	1.53 (1.37)	1.32 (1)	1.53 (1.24)	1.11 (1)
Cholecystitis	1.32 (1.17)	1.64 (1.4)	1.51 (1.27)	1.62 (1.44)
Cholangitis	3.31 (2.89)	5.37 (4.24)	3.94 (2.8)	3.07 (2.49)
Cholecystectomy	2.59 (2.39)	3.1 (2.65)	2.96 (2.56)	2.4 (2.1)
ERCP	3.82 (3.49)	5.8 (4.88)	4.44 (3.67)	3.24 (2.82)

Larger E-values indicate greater robustness to unmeasured or residual confounding. The inverse associations (cholangitis, cholecystectomy, ERCP) carry large E-values; the stone associations carry small E-values.

Discussion

In this large propensity-matched real-world cohort study, GLP-1 RA use among adults with T2DM or obesity was associated with slightly higher cumulative odds of stone-related biliary diagnoses, including cholelithiasis and choledocholithiasis. However, these findings did not translate into a higher time-adjusted rate of stone formation or a higher burden of clinically significant biliary disease. Cox hazard ratios were below 1 for all outcomes, including the stone outcomes, and Kaplan-Meier analysis showed nearly identical five-year cumulative incidence of cholelithiasis between GLP-1 RA users and matched non-users. In contrast, cholecystitis, cholangitis, cholecystectomy, and ERCP were less frequent among GLP-1 RA users. The phenotype analysis further showed that the modest stone signal was confined to patients with obesity without diabetes and was absent in patients with T2DM alone or those with both T2DM and obesity. Taken together, these findings suggest that any increase in gallstone occurrence with GLP-1 RA therapy is modest, phenotype-dependent, and does not appear to carry forward into higher rates of inflammatory, infectious, or procedural biliary complications.

The distinction between cumulative occurrence and time-to-event analysis is central to interpreting these results. Odds ratios describe how often a diagnosis was recorded during the follow-up window, whereas Cox hazard ratios and Kaplan-Meier estimates account for event timing and censoring. In our analysis, cholelithiasis and choledocholithiasis had odds ratios slightly above 1, but hazard ratios below 1. This pattern suggests that GLP-1 RA users accumulated slightly more recorded stone diagnoses over five years, but did not develop stones at a faster rate once timing was considered. One possible explanation is differential surveillance or detection, as patients receiving GLP-1 RAs may have more frequent healthcare contact, medication monitoring, laboratory testing, or abdominal imaging. Therefore, the stone signal should be interpreted as a modest increase in recorded gallstone occurrence rather than definitive evidence of accelerated stone formation.

The phenotype subgroup analysis provides further context. Higher odds of cholelithiasis and choledocholithiasis were observed only among patients with obesity without diabetes, while these associations were absent in patients with T2DM alone and in those with both T2DM and obesity. This suggests that the observed stone signal may be more closely related to obesity and weight-loss biology than to diabetes itself. Obesity is a well-established risk factor for gallstone disease, and weight loss can increase cholesterol saturation of bile and impair gallbladder emptying, both of which favor stone formation [4, 6–7, 11–12]. GLP-1 RAs and dual incretin therapies have demonstrated clinically meaningful weight loss in major obesity trials, providing a biologically plausible context for the obesity-specific signal observed in our study [13–17]. However, because individual weight trajectories, the rate of weight loss, medication dose, duration of therapy, and adherence were not available, this mechanism remains an interpretation rather than a direct finding.

Several mechanisms may explain why GLP-1 RAs have been associated with gallstone-related outcomes in prior studies. GLP-1 signaling may reduce gallbladder contractility, alter gastrointestinal motility, and impair cholecystokinin-mediated gallbladder emptying, promoting bile stasis [4–5, 18–19]. In parallel, rapid or substantial weight reduction can lead to cholesterol supersaturation of bile and gallstone formation [6–7, 11]. These mechanisms support the modest increase in recorded cholelithiasis and choledocholithiasis observed in our pooled analysis, particularly in the obesity-only subgroup. However, the absence of a higher time-adjusted event rate and the nearly identical five-year cumulative incidence of cholelithiasis suggest that these mechanisms may not translate into a clinically meaningful increase in new stone formation across the broader T2DM-or-obesity population.

A key strength of the current analysis is that it separates stone occurrence from clinically significant biliary disease. Cholelithiasis and choledocholithiasis reflect stone formation or stone migration, whereas cholecystitis and cholangitis require additional inflammatory, obstructive, infectious, or host-response factors. Cholecystectomy and ERCP represent procedural outcomes and may serve as markers of biliary disease severe enough to require intervention. In our study, the slightly higher odds of stone diagnoses did not correspond to higher odds or hazards of cholecystitis, cholangitis, cholecystectomy, or ERCP. This distinction is clinically important because asymptomatic or incidentally detected gallstones do not carry the same implications as inflammatory complications, biliary infection, or invasive procedures.

The lower odds of cholecystitis, cholangitis, cholecystectomy, and ERCP should be interpreted cautiously. These findings should not be understood as evidence that GLP-1 RAs prevent inflammatory or procedural biliary disease. Several non-causal explanations are possible, including residual confounding, prescribing patterns, differences in healthcare utilization, earlier recognition of biliary symptoms, and confounding by indication. Patients with active biliary symptoms or those approaching a biliary procedure may be less likely to initiate or continue elective metabolic therapy, which could lower inflammatory or procedural event counts in the treated group for reasons unrelated to a biological treatment effect.

At the same time, it remains biologically plausible, though unproven, that improved glycemic control, weight reduction, and anti-inflammatory effects of GLP-1-based therapies may influence progression from gallstone formation to inflammatory or infectious complications. Reviews have described anti-inflammatory actions of GLP-1 RAs beyond glucose and weight regulation [8–9]. In addition, Han et al reported a reduced risk of serious infections across several organ systems among GLP-1 RA users, although this evidence was not specific to biliary infection [20]. Thus, the lower rates of cholangitis and procedural outcomes should be viewed as hypothesis-generating associations rather than causal evidence of benefit.

Our results are consistent with several prior studies but also refine the existing literature. In a systematic review and meta-analysis of 76 randomized clinical trials, He et al found that GLP-1 RA use was associated with increased risk of gallbladder and biliary diseases, including cholelithiasis and cholecystitis, with stronger associations at higher doses, longer treatment durations, and in trials focused on weight loss [7]. Nreu et al similarly reported an association between GLP-1 RA treatment and cholelithiasis in randomized trial data [19]. Real-world pharmacovigilance and observational studies have also raised concern for gallbladder-related adverse events, including cholelithiasis and cholecystitis, although such designs are vulnerable to reporting bias and incomplete clinical confirmation [4, 11, 21–23].

There are also studies that are relevant to our interpretation. Chiang et al analyzed 55 randomized trials including 106,395 participants and found that GLP-1 RAs increased the risk of cholelithiasis but probably had little or no effect on other gastrointestinal or biliary events [24]. This closely parallels our central finding: stone diagnoses may be modestly higher, but this does not necessarily translate into more clinically significant biliary outcomes. Niu et al, using a TriNetX-based cohort of patients with T2DM, found no significant differences in cholecystitis or cholecystectomy rates between GLP-1 RA users and oral antidiabetic medication controls [25]. This supports our finding that inflammatory and procedural biliary outcomes were not increased. In contrast, Eldesouki et al, also using a TriNetX-based design, reported higher rates of cholelithiasis/choledocholithiasis, cholecystitis, and cholecystectomy at three years among patients with T2DM receiving GLP-1 RAs, while ERCP rates were not significantly different [26]. Differences between their findings and ours may reflect variation in comparator selection, BMI adjustment, index date definition, outcome definitions, follow-up duration, and analytic approach. In particular, the present analysis included BMI in the propensity model, separated metabolic phenotypes, and incorporated Cox hazard ratios and Kaplan-Meier survival analysis in addition to cumulative odds ratios.

The residual BMI imbalance deserves particular attention. Although BMI was included in the propensity score model, GLP-1 RA users remained heavier after matching. Because higher BMI is independently associated with gallstone disease, residual confounding by body weight would be expected to push biliary risk estimates upward rather than downward [12]. This means the modestly elevated stone associations may be fragile and could be partly explained by remaining differences in body weight. The E-value analysis supports this interpretation: the E-values for stone outcomes were small, suggesting

that relatively modest unmeasured or residual confounding could explain these associations. In contrast, the inverse associations for cholangitis, cholecystectomy, and ERCP had larger E-values, suggesting that they would require stronger confounding to be fully explained away. However, E-values do not prove causality and do not remove the possibility of bias; they only quantify how strong such bias would need to be.

These findings have practical clinical implications. GLP-1 RAs and dual incretin therapies provide substantial benefits for glycemic control, weight reduction, and cardiometabolic risk reduction [1–3, 13–17, 27]. Our results do not support withholding GLP-1 RA therapy solely because of concern for biliary disease. Instead, they support individualized counseling and symptom-based monitoring, particularly among patients with obesity without diabetes or those experiencing rapid weight loss. Patients should be advised to report right upper quadrant pain, postprandial abdominal pain, nausea, vomiting, jaundice, dark urine, pale stools, fever, or chills. In symptomatic patients, clinicians should consider liver biochemical testing, abdominal imaging, and gastroenterology or surgical referral when clinically appropriate. At the same time, the lower observed rates of cholecystitis, cholangitis, cholecystectomy, and ERCP should not be presented to patients as evidence that GLP-1 RAs are protective against clinically significant biliary disease.

From a pharmacovigilance perspective, these findings support continued monitoring of biliary outcomes as GLP-1 RA and dual incretin use expands across diabetes, obesity, and cardiometabolic indications. Post-marketing data and real-world studies are particularly important because biliary complications are relatively uncommon, may depend on duration of therapy, and may differ by underlying metabolic phenotype. Pharmacovigilance and meta-analytic studies have continued to identify gallbladder and biliary outcomes as safety signals requiring careful interpretation [11, 22–23]. Future studies should incorporate active comparators, longitudinal BMI and weight-loss measures, medication dose and duration, and standardized definitions of symptomatic versus incidental gallstone disease. Such designs would help clarify whether gallstone diagnoses are related to the drug itself, weight-loss biology, surveillance patterns, or baseline metabolic risk.

This study has several limitations. First, it is observational and cannot establish causality. Although propensity score matching included 20 covariates, including BMI, residual confounding remained, most notably the persistent BMI imbalance after matching. Second, confounding by indication is possible because patients with active biliary symptoms, frailty, or impending procedures may be less likely to start or continue GLP-1 RA therapy. Third, surveillance bias may have influenced the results because GLP-1 RA users may have more frequent healthcare encounters, which could increase detection and documentation of gallstones. Fourth, medication exposure was based on electronic health record documentation and may not capture dose, duration, adherence, discontinuation, or treatment interruptions. Fifth, the database did not provide individual weight trajectories or rate of weight loss, preventing direct assessment of whether weight change mediated gallstone risk. Finally, outcomes were identified using diagnosis and procedure codes without chart-level confirmation, which may introduce misclassification. Despite these limitations, the large matched cohort, inclusion of BMI in the propensity

model, time-to-event analysis, phenotype subgroups, and E-value sensitivity analysis strengthen the interpretation that any excess gallstone occurrence is modest and does not appear to translate into more clinically significant biliary disease.

Declarations

Ethics approval

This study was conducted using de-identified, aggregate electronic health record data from the TriNetX Global Collaborative Network. Because only de-identified and aggregated data were used, this study was exempt from formal institutional review board approval. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent to participate

The requirement for informed consent was waived because this study used de-identified, aggregate data and did not involve direct interaction with human participants.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no funding for this study.

Author Contribution

Chidera N Onwuzo, Huda Jaffar, Hope Marshall, Chibuike Maduka and Alvina contributed to the conception and drafting of the manuscript and preparation of the tables and figures with the data. Abdelkader Chaar and Mohamed Eisa provided critical revisions and supervision. All authors meet ICMJE authorship criteria and approve the final manuscript.

References

1. Sattar N, Lee MMY, Kristensen SL et al (2021) Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of randomised trials. Lancet Diabetes Endocrinol 9:653–662. https://doi.org/10.1016/S2213-8587(21)00203-5

2. Marso SP, Daniels GH, Brown-Frandsen K et al (2016) Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med 375:311–322. https://doi.org/10.1056/NEJMoa1603827
3. Marso SP, Bain SC, Consoli A et al (2016) Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 375:1834–1844. https://doi.org/10.1056/NEJMoa1607141
4. Ramírez-Mejía MM, Ponciano-Rodriguez G, Eslam M, Méndez-Sánchez N (2025) GLP-1 receptor agonists and gallbladder disease risk: insights into molecular mechanisms and clinical implications. Ther Adv Endocrinol Metab 16:20420188251406456. https://doi.org/10.1177/20420188251406456
5. Shaddinger BC, Young MA, Billiard J, Collins DA, Hussaini A, Nino A (2017) Effect of albiglutide on cholecystokinin-induced gallbladder emptying in healthy individuals: a randomized crossover study. J Clin Pharmacol 57:1322–1329. https://doi.org/10.1002/jcph.940
6. Gameil MA, Yousef EAAM, Marzouk RE, Emara MH, Abdelkader AH, Salama RI (2024) The relative risk of clinically relevant cholelithiasis among glucagon-like peptide-1 receptor agonists in patients with type 2 diabetes mellitus, real-world study. Diabetol Metab Syndr 16:293. https://doi.org/10.1186/s13098-024-01526-2
7. He L, Wang J, Ping F et al (2022) 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 182:513–519. https://doi.org/10.1001/jamainternmed.2022.0338
8. Alharbi SH (2024) Anti-inflammatory role of glucagon-like peptide 1 receptor agonists and its clinical implications. Ther Adv Endocrinol Metab 15:20420188231222367. https://doi.org/10.1177/20420188231222367
9. Wong CK, Drucker DJ (2025) Anti-inflammatory actions of glucagon-like peptide-1-based therapies beyond metabolic benefits. J Clin Invest 135:e194751. https://doi.org/10.1172/JCI194751
10. Ludwig RJ, Anson M, Zirpel H et al (2025) A comprehensive review of methodologies and application to use the real-world data and analytics platform TriNetX. FrontPharmacol 16:1516126. https://doi.org/10.3389/fphar.2025.1516126
11. Tao C, Zhang Y, Wan T et al (2025) Glucagon-like peptide-1 receptor agonist-induced cholecystitis and cholelithiasis: a real-world pharmacovigilance analysis using the FAERS database. Front Pharmacol 16:1557691. https://doi.org/10.3389/fphar.2025.1557691
12. Yuan S, Gill D, Giovannucci EL, Larsson SC (2022) Obesity, type 2 diabetes, lifestyle factors, and risk of gallstone disease: a Mendelian randomization investigation. ClinGastroenterol Hepatol 20:e529–e537. https://doi.org/10.1016/j.cgh.2020.12.034
13. Wilding JPH, Batterham RL, Calanna S et al (2021) Once-weekly semaglutide in adults with overweight or obesity. N Engl J Med 384:989–1002. https://doi.org/10.1056/NEJMoa2032183
14. Jastreboff AM, Aronne LJ, Ahmad NN et al (2022) Tirzepatide once weekly for the treatment of obesity. N Engl J Med 387:205–216. https://doi.org/10.1056/NEJMoa2206038
15. Rubino D, Abrahamsson N, Davies M et al (2021) Effect of continued weekly subcutaneous semaglutide vs placebo on weight loss maintenance in adults with overweight or obesity: the STEP 4 randomized clinical trial. JAMA 325:1414–1425. https://doi.org/10.1001/jama.2021.3224

16. Garvey WT, Batterham RL, Bhatta M et al (2022) Two-year effects of semaglutide in adults with overweight or obesity: the STEP 5 trial. Nat Med 28:2083–2091. https://doi.org/10.1038/s41591-022-02026-4
17. Garvey WT, Frias JP, Jastreboff AM et al (2023) Tirzepatide once weekly for the treatment of obesity in people with type 2 diabetes: SURMOUNT-2, a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet 402:613–626. https://doi.org/10.1016/S0140-6736(23)01200-X
18. Kim JA, Yoo HJ (2025) Exploring the side effects of GLP-1 receptor agonist: to ensure its optimal positioning. Diabetes Metab J 49:525–541. https://doi.org/10.4093/dmj.2025.0242
19. Nreu B, Dicembrini I, Tinti F, Mannucci E, Monami M (2020) Cholelithiasis in patients treated with glucagon-like peptide-1 receptor: an updated meta-analysis of randomized controlled trials. Diabetes Res Clin Pract 161:108087. https://doi.org/10.1016/j.diabres.2020.108087
20. Han S, Liu Y, Xing B et al (2025) Association of glucagon-like peptide-1 receptor agonist use with risk of infections: a systematic review and meta-analysis. J Infect91:106645. https://doi.org/10.1016/j.jinf.2025.106645
21. Woronow D, Chamberlain C, Niak A, Avigan M, Houstoun M, Kortepeter C (2022) Acute cholecystitis associated with the use of glucagon-like peptide-1 receptor agonists reported to the US Food and Drug Administration. JAMA Intern Med 182:1104–1106. https://doi.org/10.1001/jamainternmed.2022.3810
22. Zeng Q, Xu J, Mu X, Shi Y, Fan H, Li S (2023) Safety issues of tirzepatide: pancreatitis and gallbladder or biliary disease in type 2 diabetes and obesity: a systematic review and meta-analysis. Front Endocrinol 14:1214334. https://doi.org/10.3389/fendo.2023.1214334
23. Mishra R, Raj R, Elshimy G et al (2023) Adverse events related to tirzepatide. J Endocr Soc 7:bvad016. https://doi.org/10.1210/jendso/bvad016
24. Chiang CH, Jaroenlapnopparat A, Colak SC et al (2025) Glucagon-like peptide-1 receptor agonists and gastrointestinal adverse events: a systematic review and meta-analysis. Gastroenterology 169:1268–1281. https://doi.org/10.1053/j.gastro.2025.06.003
25. Niu C, Sun K, Zhang J et al (2025) Gastrointestinal and hepatobiliary safety of glucagon-like peptide-1 receptor agonists in patients with type 2 diabetes. Am J Gastroenterol. https://doi.org/10.14309/ajg.0000000000003760
26. Eldesouki MH, Alkasabrah O, Kloub M et al (2026) Cohort study: risk of gallstones and biliary complications with glucagon-like peptide-1 receptor agonists in type 2 diabetes. United European Gastroenterol J 14:e70238. https://doi.org/10.1002/ueg2.70238
27. Lincoff AM, Brown-Frandsen K, Colhoun HM et al (2023) Semaglutide and cardiovascular outcomes in obesity without diabetes. N Engl J Med 389:2221–2232. https://doi.org/10.1056/NEJMoa2307563

Figures

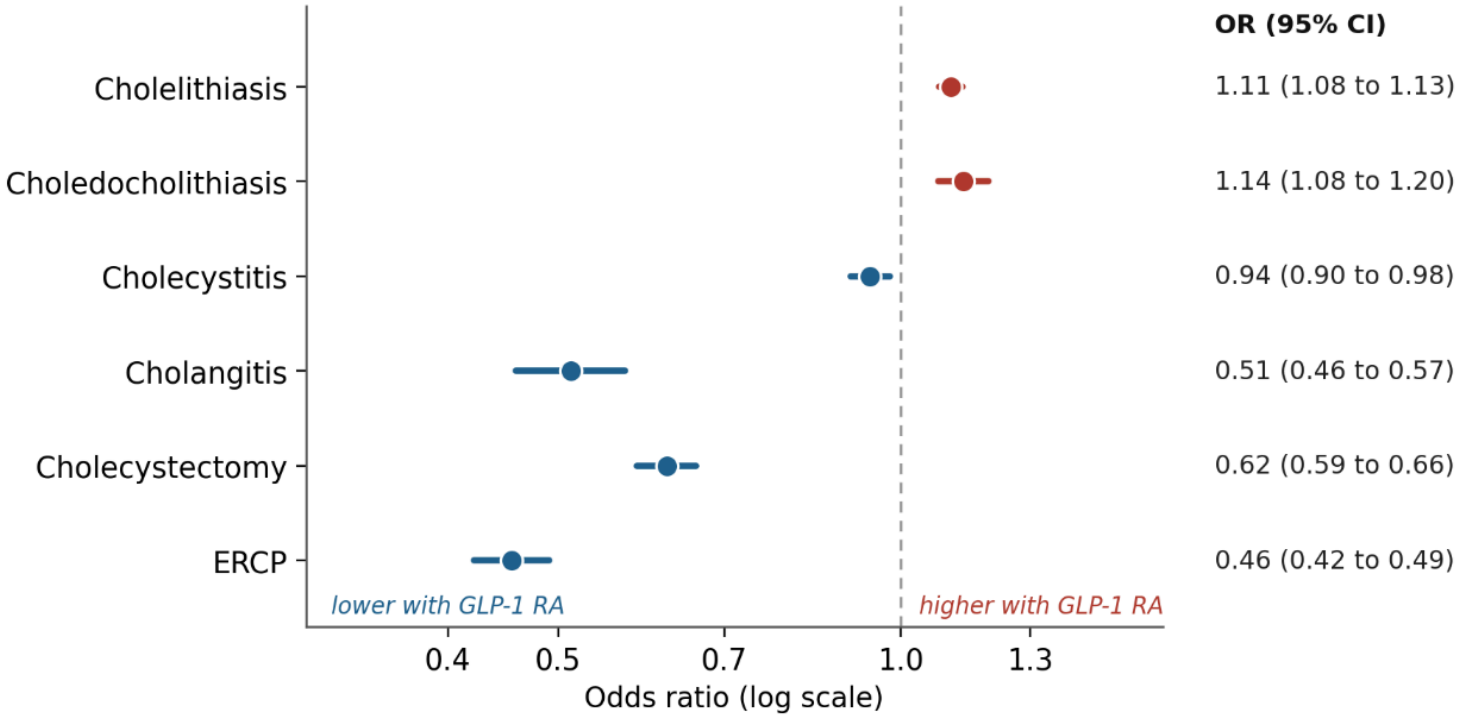


Figure 1

Odds ratios for biliary outcomes in the pooled cohort, on a log scale. Stone outcomes are slightly higher; cholecystitis, cholangitis, cholecystectomy, and ERCP are lower.

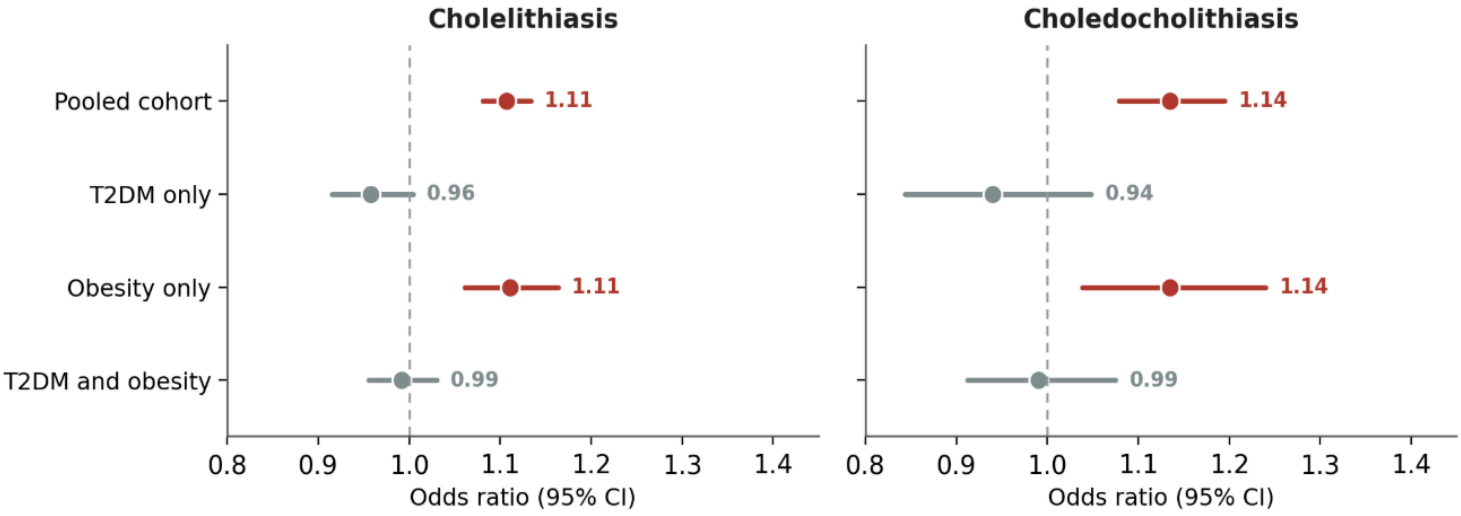


Figure 2

Stone outcomes by metabolic phenotype. The higher odds are confined to obesity without diabetes and are absent in every group that includes diabetes.

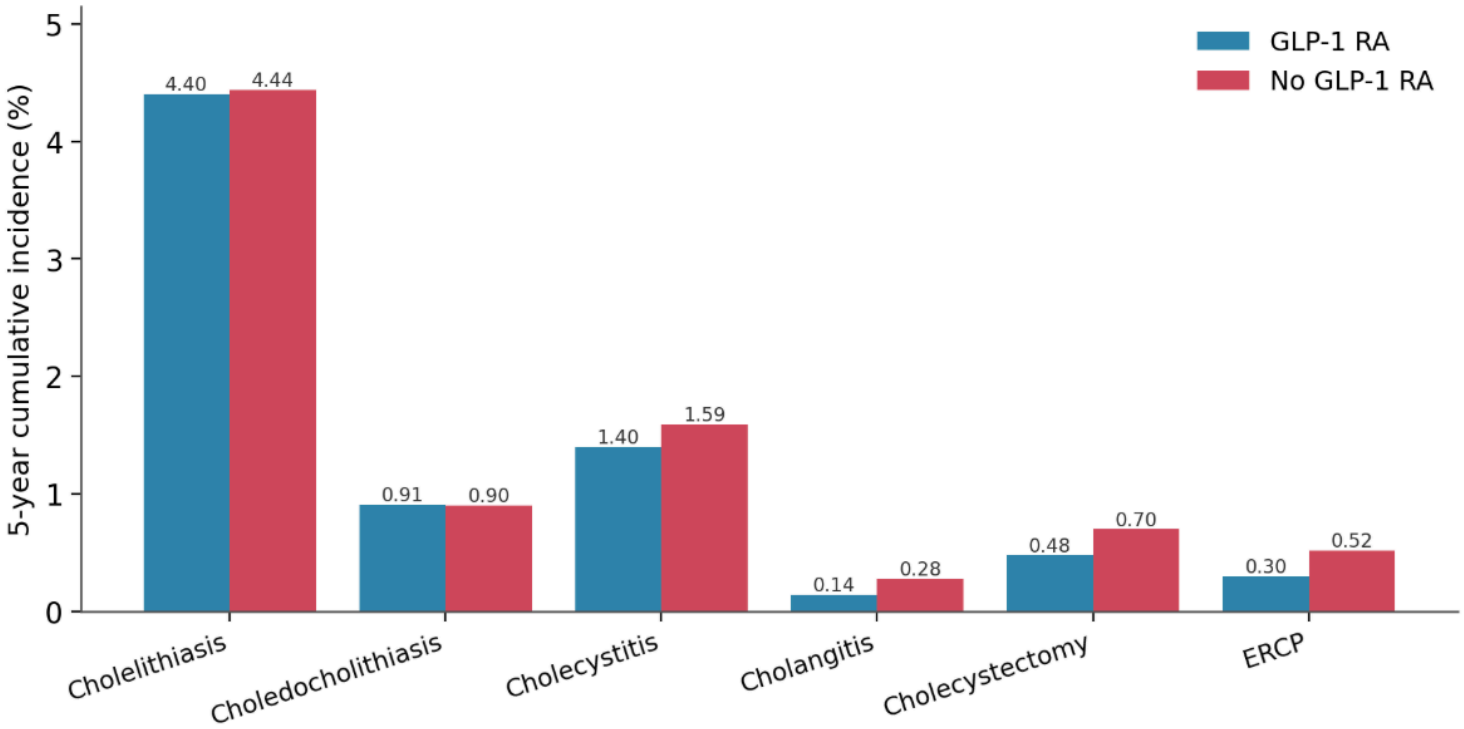


Figure 3

Five-year cumulative incidence by cohort, from Kaplan-Meier survival estimates. Cholelithiasis incidence is almost identical between groups, while cholecystitis, cholangitis, cholecystectomy, and ERCP are lower with GLP-1 RA.