OPEN ACCESS

Original research

Glucagon-like peptide-1 receptor agonists and risk of major adverse liver outcomes in patients with chronic liver disease and type 2 diabetes

Axel Wester ,[1] Ying Shang ,[1] Emilie Toresson Grip,[1,2] Anthony A Matthews,[3] Hannes Hagström [1,4]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2023-330962).

[1]Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden
[2]Quantify Research, Stockholm, Sweden
[3]Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[4]Division of Hepatology, Department of Upper GI, Karolinska University Hospital, Stockholm, Sweden

Correspondence to
Dr Axel Wester, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden; axel.wester@ki.se

AAM and HH are joint senior authors.

Received 21 August 2023
Accepted 13 January 2024

To cite: Wester A, Shang Y, Toresson Grip E, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2023-330962

ABSTRACT

Objective Phase II trials suggest glucagon-like peptide-1 receptor (GLP1) agonists resolve metabolic dysfunction-associated steatohepatitis but do not affect fibrosis regression. We aimed to determine the long-term causal effect of GLP1 agonists on the risk of major adverse liver outcomes (MALO) in patients with any chronic liver disease and type 2 diabetes.

Design We used observational data from Swedish healthcare registers 2010–2020 to emulate a target trial of GLP1 agonists in eligible patients with chronic liver disease and type 2 diabetes. We used an inverse-probability weighted marginal structural model to compare parametric estimates of 10-year MALO risk (decompensated cirrhosis, hepatocellular carcinoma, liver transplantation or MALO-related death) in initiators of GLP1 agonists with non-initiators. We randomly sampled 5% of the non-initiators to increase computational efficiency.

Results GLP1 agonist initiators had a 10-year risk of MALO at 13.3% (42/1026) vs 14.6% in non-initiators (1079/15 633) in intention-to-treat analysis (risk ratio (RR)=0.91, 95% CI=0.50 to 1.32). The corresponding 10-year per-protocol risk estimates were 7.4% (22/1026) and 14.4% (1079/15 633), respectively (RR=0.51, 95% CI=0.14 to 0.88). The per-protocol risk estimates at 6 years were 5.4% (21/1026) vs 9.0% (933/15 633) (RR=0.60, 95% CI=0.29 to 0.90) and at 8 years 7.2% (22/1026) vs 11.7% (1036/15 633) (RR=0.61, 95% CI=0.21 to 1.01).

Conclusion In patients with chronic liver disease and type 2 diabetes who adhered to therapy over time, GLP1 agonists may result in lower risk of MALO. This suggests that GLP1 agonists are promising agents to reduce risk of chronic liver disease progression in patients with concurrent type 2 diabetes, although this needs to be corroborated in randomised trials.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Glucagon-like peptide-1 receptor (GLP1) agonists might resolve metabolic dysfunction-associated steatohepatitis, but their effect on hard clinical outcomes in patients with chronic liver diseases of any aetiology and concurrent type 2 diabetes is unknown.

WHAT THIS STUDY ADDS

⇒ Using Swedish register data, we emulated a target trial of GLP1 agonists in patients with chronic liver disease and type 2 diabetes and fitted an inverse-probability weighted marginal structural model to estimate 10-year risks of major adverse liver outcomes (MALO). The risk of MALO was 49% lower in initiators of GLP1 agonists in the per-protocol analysis, but our data were not compatible with a protective intention-to-treat effect.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ GLP1 agonists might be a treatment option to reduce MALO risk in patients with type 2 diabetes and any chronic liver disease who adhere to the treatment over time, although this would need to be corroborated by randomised clinical trials.

INTRODUCTION

Chronic liver diseases are highly prevalent and can progress to decompensated cirrhosis, hepatocellular carcinoma (HCC) and liver-related death.[1–5] Type 2 diabetes strongly predicts presence and severity of metabolic dysfunction-associated steatotic liver disease (MASLD) and is also a major risk factor for disease progression in other liver diseases, likely due to interaction of hepatic steatosis and steatohepatitis with other liver diseases.[6–9]

Currently, no approved pharmacotherapy exists for MASLD, but one promising drug class is glucagon-like peptide-1 receptor (GLP1) agonists, which are currently approved in patients with type 2 diabetes or obesity to achieve weight loss and control blood glucose.[10–12] Importantly, phase II trials indicate that GLP1 agonists resolve metabolic dysfunction-associated steatohepatitis (MASH) in patients with non-cirrhotic MASLD but do not cause fibrosis regression.[13 14] Large phase III trials that aim to estimate the effect of GLP1 agonists on resolving MASH or reducing hepatic fibrosis are, however, many years from completion.[15]

Although achieving surrogate histopathological endpoints (eg, fibrosis regression) is considered by regulatory agencies to likely translate to improved prognosis, robust evidence is needed to understand if GLP1 agonists reduce the risk of long-term clinical outcomes, such as liver decompensation or

HCC.[16 17] Given that the metabolic syndrome is a major driver of liver-related outcomes both in patients with MASLD and other chronic liver diseases such as alcohol-related liver disease or viral hepatitis C, there could be a similar effect of GLP1 agonists in patients with chronic liver diseases of any aetiology with concomitant metabolic traits, such as type 2 diabetes.[18–20] For instance, insulin resistance is the strongest predictor of liver fibrosis in patients with alcohol-related liver disease.[21] Therefore, we designed a target trial that would estimate the long-term causal effect of GLP1 agonists on major adverse liver outcomes (MALO) in patients with any chronic liver disease and type 2 diabetes, and then emulated it using observational data from Swedish healthcare registers.

METHODS

Data sources

The Decoding the Epidemiology of LIVER disease in Sweden (DELIVER) cohort includes data from Swedish national healthcare registers on all patients with any chronic liver disease in Sweden 1964–2020.[22] The data include all International Classification of Diseases (ICD) codes from inpatient and specialised outpatient care, dates and causes of death, and automatically recorded information on filled prescriptions from any pharmacy in Sweden.[23–29] The positive predictive value is 96% for MASLD with comorbid type 2 diabetes, and >90% for most diagnoses related to cirrhosis.[30 31] A detailed overview of the registers is provided in online supplemental methods.

Target trial specification and emulation

A causal question is best answered by data from randomised trials, but when unavailable, researchers often resort to observational data from existing databases. To avoid common methodological pitfalls in observational studies, causal inference from such data can be viewed as an attempt to emulate a hypothetical pragmatic randomised trial—a target trial.[32] An overview of the target trial emulation concept is provided in reference.[33] After specifying the target trial protocol, it is emulated using the available observational data and appropriate methodology. For this observational study, we first specified the protocol of a target trial that would estimate the effect of GLP1 agonists on MALO risk in patients with chronic liver disease and type 2 diabetes. We then emulated the target trial using data from DELIVER. Table 1 summarises all protocol components from the target trial and its emulation, which we describe in detail below. All diagnoses and medications were defined by the ICD or Anatomical Therapeutic Chemical (ATC) codes in online supplemental tables 1–3. Diagnoses were identified from January 1997 and forward (when ICD-10 was introduced in Sweden), and drugs from July 2005 and forward (when the Swedish Prescribed Drug Register was initiated).

Eligibility criteria

All Swedish residents ≥18 years of age between January 2010, when uptake of GLP1 agonists increased in Sweden, and November 2020 with any chronic liver disease and type 2 diabetes were identified.[22] To avoid structural positivity issues (ie, patients with zero probability of initiating a GLP1 agonist at baseline), patients were required to have filled at least one prescription of metformin within a year before baseline corresponding to a daily dose of ≥1 g (ie, patients potentially eligible for second-line treatment with a GLP1 agonist). Patients were excluded if they previously filled prescriptions of GLP1 agonists, had a history of a contraindication to GLP1 agonists (defined as pancreatitis, inflammatory bowel disease or severe chronic kidney disease) or prior MALO (defined below). As the non-initiator population was large, we randomly sampled 5% of this group to increase computational efficiency (figure 1).

Treatment strategies

We compared two treatment strategies: initiation of any GLP1 agonist (ATC code A10BJ) at baseline and continuation of treatment over follow-up unless a contraindication was diagnosed after baseline; and no initiation of a GLP1 agonist at baseline and continuation of no GLP1 agonist treatment during follow-up, unless indicated as deemed by the treating physician. Since we lacked data to specifically determine who were indicated for GLP1 agonists, we assumed that it was indicated in all non-initiators who started the drug during follow-up. We assessed drug continuation by summing the number of months of filled prescriptions. A gap between two successive prescriptions was allowed if it was less than twice the time intended for the most recently filled prescription. For example, an initiator who filled a prescription for 3 months treatment was considered to have stopped the treatment after 6 months, unless the prescription was refilled before that.

Treatment assignment

Patients were classified into two groups according to the strategy their data were compatible with at baseline, that is, GLP1 agonist initiators and non-initiators. We assumed groups were exchangeable at baseline conditional on baseline covariates (similar probability of initiating the drug in both arms, within levels of the covariates): age, sex, education (<10, 10–12 and >12 years), diabetes duration, liver disease aetiology, compensated cirrhosis, and a range of comorbidities and medications: obesity, cardiovascular disease, microvascular complications to diabetes, chronic obstructive pulmonary disease, alcohol use disorder, mental health disorder, the use of antidiabetic medications except metformin or GLP1 agonists and direct-acting antivirals in patients with viral hepatitis. As the relationship between age and the probability of initiating a GLP1 agonist might not be linear, we modelled age using linear, quadratic and cubic terms. If patients had coding for more than one liver disease aetiology, they were classified according to a predefined hierarchy (online supplemental methods).

Outcome

The outcome of interest was the first MALO during follow-up, defined as decompensated cirrhosis (variceal bleeding, ascites, portal hypertension or hepatorenal syndrome), HCC, liver transplantation or MALO-related death. MALO was defined by ICD codes in the National Patient Register (main or secondary diagnosis), the Cancer Register or the Cause of Death Register (main or contributing cause) (online supplemental table 3). These outcomes have been validated and found to have positive predictive values >90% (online supplemental table 4).[31]

Follow-up

Everyone was followed from baseline to the earliest of MALO, emigration from Sweden, 10 years of follow-up or December 2020. The follow-up was measured in calendar months.

Causal contrasts

We estimated observational analogues of the intention-to-treat and per-protocol effects.

Table 1 Protocol of a target trial and an emulated trial using observational data

Protocol component	Target trial	Emulated trial using observational data
Eligibility criteria	Inclusion Swedish residents ≥18 years of age between January 2010 and November 2020 with any chronic liver disease and type 2 diabetes previously diagnosed at any point, who are currently using at least 1 g of metformin per day and have used metformin for at least 6 months. Exclusion Any previously filled prescription of GLP1 agonists. History of any contraindications to GLP1 agonists (pancreatitis, inflammatory bowel disease or severe chronic kidney disease). History of major adverse liver outcomes.	Same as the target trial. Patients are required to have filled at least one prescription of metformin the last year, corresponding to a daily dose of 1 g.
Treatment strategies	1. Initiation of a GLP1 agonist at baseline and continued treatment during the follow-up, unless contraindicated. 2. No initiation of a GLP1 agonist at baseline and continuation of no treatment during the follow-up, unless indicated.	Same as the target trial.
Treatment assignment	Random unblinded assignment at baseline.	Patients are classified as initiators or non-initiators according to what their data at baseline are compatible with. Randomisation is emulated by adjusting for baseline confounders.
Outcome	Major adverse liver outcomes, which will be a composite outcome including decompensated cirrhosis (variceal bleeding, ascites, portal hypertension or hepatorenal syndrome), hepatocellular carcinoma, need for liver transplantation or major adverse liver outcome-related death.	Same as the target trial.
Follow-up	Start (baseline) Any calendar month in which all eligibility criteria are met. Patients can be enrolled in several different target trials. End The calendar month of first outcome, emigration from Sweden, 10 years of follow-up or December 2020, whichever occurs first.	Same as the target trial.
Causal contrasts	Intention-to-treat and per-protocol effects.	Observational analogues of the intention-to-treat and per-protocol effects.
Statistical analyses	Intention-to-treat analysis Estimate the risk curves in each group defined by assigned treatment strategy via a parametric pooled logistic model with an indicator for treatment group, a flexible time-varying intercept, product terms between treatment group and time, and a trial indicator. Per-protocol analysis Same as above, but individuals will be censored when they deviate from their assigned strategy and inverse-probability weights will be applied to adjust for baseline and time-varying covariates associated with adherence. The effect of initiating GLP1 agonists will be studied in prespecified subgroups according to liver disease aetiology (MASLD, other than MASLD) and liver disease severity at baseline (compensated cirrhosis, no cirrhosis).	Same as the target trial with sequential emulation (starting in each calendar month of the study period) and additional adjustments for baseline covariates associated with treatment initiation in both the intention-to-treat and per-protocol analyses.

.GLP1, glucagon-like peptide-1 receptor; MASLD, metabolic dysfunction-associated steatotic liver disease.

Statistical analyses

We sequentially emulated the target trial as a series of separate target trials starting in each 131 calendar months between January 2010 and November 2020, meaning that patients could enter multiple target trials if eligible. To avoid immortal time bias, the baseline is best defined as the time when eligible patients initiate a treatment strategy.[32] The GLP agonist initiators naturally have one such time point. The definition of the baseline is, however, more challenging in the non-initiators since the same individual can be eligible at multiple times. One solution that avoids immortal time bias is to emulate a target trial that uses all those eligibility times as the baseline and consider

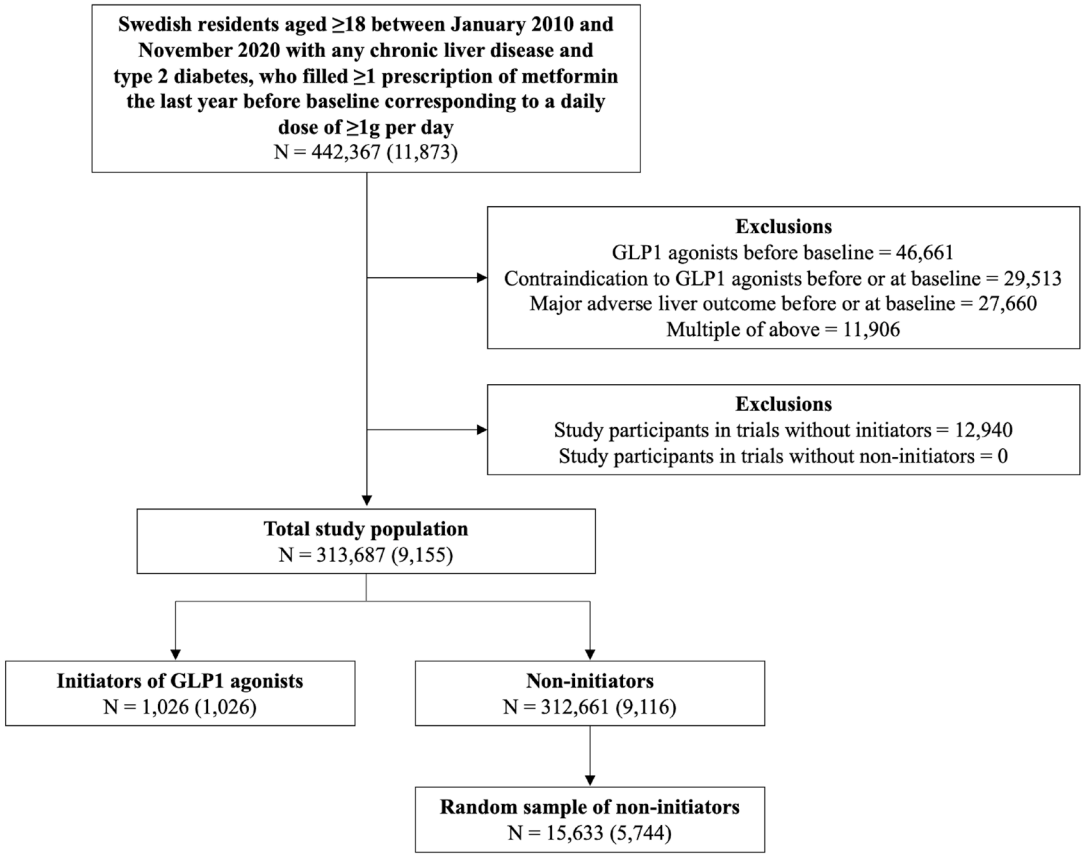

Figure 1 Flow chart of the study population. Numbers represent study participants (initiators or non-initiators), while numbers in parentheses represent the corresponding number of unique patients. Note that the numbers of excluded study participants represent the total number of times that unique patients were non-eligible for any of the emulated target trials. If somebody was non-eligible for all the 131 emulated target trials, they would contribute with an addition of 131 to this total number.

each individual at each of those times as different individuals.[34] For example, a patient who fulfilled all eligibility criteria in our study in January 2010 and did not initiate a GLP1 agonist in that month would enter the target trial that started in January 2010 as a non-initiator. If still eligible in February 2010, this patient would also enter the target trial that started in this month as a non-initiator. If the same patient initiated a GLP1 agonist in March 2010 and still fulfilled all eligibility criteria, the patient would enter this target trial as an initiator and be non-eligible for all subsequent target trials as previous use of GLP1 agonists was an exclusion criterion. Target trials with only initiators or only non-initiators were excluded. Allowing repeated eligibility is statistically more efficient than choosing only 1 month as baseline and accounts for the fact that patients can be eligible in several different months during the study period.[35 36]

We fitted a marginal structural model using parametric pooled logistic regression with an indicator for treatment group, a flexible time-varying intercept (linear and quadratic terms), product terms between treatment group and time, and a target trial indicator to pool data for all the emulated trials and estimate intention-to-treat and per-protocol effects. For estimation of the intention-to-treat effect, we weighted the model using inverse-probability of treatment weights (IPTW). The IPTW models included all baseline covariates, and the weights were stabilised. The balance between treatment groups was assessed using standardised mean differences (SMD) and inspection of kernel density plots. An SMD<0.1 is generally regarded to indicate good balance.[37] For estimation of the per-protocol effect, the same marginal structural model as above was used, but patients were additionally censored when deviating from their assigned treatment strategy and stabilised inverse-probability of censoring weights (IPCWs) were applied to adjust for baseline and time-varying covariates associated with adherence.[34 38] The marginal structural model was weighted using the product of the IPTWs and the IPCWs. The IPCW models included age, sex, education, diabetes duration and liver disease aetiology at baseline, and the following time-varying covariates: compensated cirrhosis and the same range of comorbidities and medications as described above. If patients stopped GLP1 agonist treatment because of a contraindication during follow-up (eg, an episode of pancreatitis), or started GLP1 agonist treatment, their censoring weights remained constant from that date forward. The weights are described in online supplemental table 5.

The average 10-year absolute risks under each strategy were estimated using the predicted values from the marginal structural models, then resulting risk differences (RDs) and risk ratios (RRs) were calculated. Non-parametric bootstrapping with 500 replications was used to estimate 95% CIs.

We examined the intention-to-treat effect in subgroups according to liver disease aetiology (MASLD, other than MASLD), and liver disease severity at baseline (compensated cirrhosis, no cirrhosis). Several sensitivity analyses were done to assess the robustness of our results. First, we updated the eligibility criteria to include those with a lower daily dose of

metformin (0.5 g instead of 1 g). Second, the time gap between two successive prescriptions was restricted to ≤30 days. Third, we censored non-initiators if they initiated a GLP1 agonist during follow-up, regardless of whether it was indicated, and IPCWs were applied as described above. Fourth, the inverse-probability weights were truncated at the 1st and 99th percentile before being applied to the marginal structural model, to avoid the impact of extreme values on the risk estimates. Fifth, we used standardisation to adjust for confounding at baseline, rather than IPTW.[39] In this analysis, the IPCWs were used in the per-protocol analysis as described above. Sixth, we estimated intention-to-treat and per-protocol point estimates including all non-initiators (without sampling 5%), to assess whether the sampling affected our risk estimates. Finally, to estimate how strongly an unmeasured confounder would need to be associated with both the exposure and outcome to fully explain any differences in risk between treatment groups, we calculated the E-value.[40]

We additionally calculated the intention-to-treat and per-protocol effects at 2, 4, 6 and 8 years of follow-up. Because 2 years is a more plausible duration of a future randomised trial of GLP1 agonists in patients with chronic liver disease than 10 years, we computed the minimum sample size required in a clinical trial of GLP1 agonists to demonstrate an effect of equal strength as our estimated 2-year RR, using a 5% alpha and 80% power.

Analyses were done in Stata V.17.0 (StataCorp).

Patient and public involvement

Patients or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

RESULTS

We included 1026 initiators of GLP1 agonists and 15 633 non-initiators, who participated in up to 123 target trials. The patient characteristics at baseline before and after weighting are summarised in table 2. After IPTW, all baseline characteristics were well balanced (SMD<0.1). Kernel density plots also indicated good balance (online supplemental figure 1). Of the initiators, 635 (61.9%) started treatment with liraglutide, 231 (22.5%) semaglutide, 120 (11.7%) dulaglutide, 25 (2.4%) exenatide and 15 (1.5%) lixisenatide at baseline.

In the intention-to-treat analysis, initiators and non-initiators were followed for a median (p25–p75) of 64 (36–96) and 76 months (50–100), respectively. 350 of 605 (57.9%) initiators still at risk at 2 years were continuous users at this time. The corresponding numbers at 4, 6, 8 and 10 years were 143 of 331 (43.2%), 64 of 188 (34.0%), 24 of 98 (24.5%) and 5 of 19 (26.3%), respectively. In the per-protocol analysis, where the patients were censored if they deviated from their assigned treatment strategy, the median follow-up was 43 (21–75) and 76 months (50–100), respectively. After a median of 14 months (8–26) follow-up, 517 (50.4%) initiators stopped the treatment; 21 (2.1%) initiators stopped the treatment after developing a contraindication, and 496 (48.3%) stopped the treatment without one of the prespecified contraindications. Of the 1026 initiators, 361 (35.2%) were censored in the per-protocol analysis because they stopped the treatment with a prespecified contraindication the first 2 years, another 92 (9.0%) between 2 and 4 years, 28 (2.7%) between 4 and 6 years, 11 (1.1%) between 6 and 8 years, and 4 (0.4%) between 8 and 10 years. Among the non-initiators, 2357 (15.1%) started treatment with a GLP1 agonist after a median follow-up of 31 months (14–52).

In the intention-to-treat analysis, MALO occurred in 42 initiators and 1079 non-initiators. The events in non-initiators corresponded to 486 distinct events (since participation in multiple target trials was allowed, some contributed with events to more than one target trial). The 10-year risk of MALO was 13.3% (95% CI=7.4% to 19.2%) in the initiators and 14.6% (95% CI=13.1% to 16.1%) in the non-initiators (RD=−1.3, 95% CI=−7.2 to 4.6, RR=0.91, 95% CI=0.50 to 1.32) (table 3, figure 2A).

In per-protocol analysis, we observed 22 events of MALO in the initiators and 1079 in the non-initiators. The 10-year MALO risk was 7.4% (95% CI=2.1% to 12.6%) in the initiators and 14.4% (95% CI=12.9% to 15.9%) in the non-initiators (RD=−7.1, 95% CI=−12.5 to −1.6; RR=0.51, 95% CI=0.14 to 0.88) (table 3, figure 2B). This corresponds to a number needed to treat (initiate and continue with the GLP1 agonist treatment strategy) to avoid one event of MALO over the course of 10 years of 14 (95% CI=8 to 63).

Results from the intention-to-treat subgroup analyses are summarised in table 4. In patients with MASLD, the 10-year risk of MALO was 15.8% in the initiators and 11.2% in the non-initiators (RR=1.41, 95% CI=0.53 to 2.30). The 10-year risk in patients with compensated cirrhosis was 36.5% and 34.6% in initiators and non-initiators, respectively (RR=1.05, 95% CI=0.20 to 1.91).

The risk estimates from a sensitivity analysis that allowed patients to have a lower dose of metformin at baseline were similar to the main analyses (intention-to-treat RR=0.94, 95% CI=0.58 to 1.30; per-protocol RR=0.63, 95% CI=0.27 to 0.99) (online supplemental table 6). The per-protocol estimates were also similar when restricting the gap between two successive prescriptions to ≤30 days (RR=0.48, 95% CI=0.11 to 0.85) (online supplemental table 7). In addition, similar per-protocol estimates were found when the non-initiators were censored if they initiated a GLP1 agonist during follow-up (RR=0.55, 95% CI=0.18 to 0.92) (online supplemental table 8). Both the intention-to-treat and per-protocol estimates were similar to the main analyses when the weights were truncated (intention-to-treat RR=0.84, 95% CI=0.48 to 1.20; per-protocol RR=0.55, 95% CI=0.16 to 0.93) (online supplemental table 9). Moreover, results were similar to the main analysis when using standardisation to adjust for confounders at baseline (intention-to-treat RR=0.87, 95% CI=0.53 to 1.22; per-protocol RR=0.58, 95% CI=0.19 to 0.97) (online supplemental table 10). Point estimates were similar to the main analysis when including all non-initiators (n=312 661) (intention-to-treat RR=1.01, per-protocol RR=0.53) (online supplemental table 11). The E-value for the RRs in the main per-protocol analysis was 3.33 for the point estimate and 1.53 for the 95% CI, suggesting how strong an unmeasured confounder needs to be to fully explain the estimated per-protocol effect and shift the 95% CI to include the null.

Risk estimates for the intention-to-treat and per-protocol analyses at 2, 4, 6 and 8 years are presented in table 3. At 2 years follow-up, a more plausible duration of a clinical trial than 10 years, the risk of MALO was 2.3% in the initiators and 3.2% in the non-initiators (overall event probability of 2.7%) when analysed by the intention-to-treat principle (RD=−0.9, 95% CI=−2.2 to 0.4; RR=0.72, 95% CI=0.31 to 1.13). A clinical trial of GLP1 agonists on the risk of

Table 2 Patient characteristics at baseline before and after inverse-probability of treatment weighting

	Before weighting			After weighting		
	GLP1 agonist initiators (n=1026)	Non-initiators (n=15 633)	SMD	GLP1 agonist initiators (n=968)	Non-initiators (n=15 326)	SMD
Included unique patients, n	1026	5744				
Sex, n (%)			0.10			0.03
Men	577 (56.2)	9549 (61.1)		610 (63.0)	9420 (61.5)	
Women	449 (43.8)	6084 (38.9)		358 (37.0)	5906 (38.5)	
Age in years, median (p25–p75)	60 (52–66)	63 (56–71)	0.39	62 (54–70)	63 (56–70)	0.09
Country of birth, n (%)			0.02			0.03
Nordic	758 (73.9)	11 432 (73.1)		735 (75.9)	11 406 (74.4)	
Other	268 (26.1)	4201 (26.9)		233 (24.1)	3920 (25.6)	
Education in years, n (%)						
<10	281 (27.4)	5335 (34.1)	0.15	337 (34.8)	5271 (34.4)	0.01
10–12	529 (51.6)	7208 (46.1)	0.11	459 (47.4)	7263 (47.4)	0.00
>12	197 (19.2)	2779 (17.8)	0.04	172 (17.8)	2793 (18.2)	0.01
Missing	19 (1.9)	311 (2.0)	0.01	0 (0.0)	0 (0.0)	
Aetiology of liver disease, n (%)						
ALD with or without viral hepatitis	102 (9.9)	1863 (11.9)	0.04	108 (11.2)	1839 (12.0)	0.03
Viral hepatitis without ALD	340 (33.1)	6267 (40.1)	0.21	375 (38.8)	5993 (39.1)	0.01
Other	146 (14.2)	2427 (15.5)	0.06	146 (15.0)	2384 (15.6)	0.01
MASLD	438 (42.7)	5076 (32.5)	0.14	339 (35.0)	5110 (33.3)	0.04
Compensated cirrhosis	161 (15.7)	2721 (17.4)	0.05	167 (17.3)	2668 (17.4)	0.00
Diabetes duration in years, median (p25–p75)	7 (4–11)	7 (3–10)	0.11	6 (3–11)	7 (4–10)	0.00
Chronic liver disease duration in years, median (p25–p75)	7 (3–12)	8 (3–12)	0.06	7 (3–13)	8 (3–12)	0.01
Antidiabetic medication other than metformin or GLP1 agonists before baseline, n (%)	857 (83.5)	10 232 (65.5)	0.42	675 (69.7)	10 174 (66.4)	0.07
Insulin	540 (52.6)	5866 (37.5)	0.31	425 (43.9)*	5850 (38.2)*	0.11
Sodium-glucose cotransporter-2 inhibitors	207 (20.2)	1154 (7.4)	0.38	159 (16.5)*	1142 (7.4)*	0.28
Dipeptidyl Peptidase-4 inhibitors	386 (37.6)	3114 (19.9)	0.40	300 (31.0)*	3094 (20.2)*	0.25
Other	425 (41.4)	5624 (36.0)	0.11	331 (34.2)*	5537 (36.1)*	0.04
Comorbidities at baseline, n (%)						
Obesity	408 (39.8)	3270 (20.9)	0.42	232 (23.9)	3404 (22.2)	0.04
Cardiovascular disease	656 (63.9)	9481 (60.7)	0.07	589 (60.8)	9367 (61.1)	0.01
Microvascular complications to diabetes	208 (20.3)	2989 (19.1)	0.03	182 (18.8)	2944 (19.2)	0.01
Cancer other than HCC	131 (12.8)	2478 (15.9)	0.09	136 (14.0)	2427 (15.8)	0.05
Chronic obstructive pulmonary disease	82 (8.0)	1010 (6.5)	0.06	66 (6.9)	1014 (6.6)	0.01
Alcohol use disorder	121 (11.8)	2278 (14.6)	0.08	143 (14.8)	2243 (14.6)	0.00
Mental health disorder	290 (28.3)	3960 (25.3)	0.07	259 (26.7)	3935 (25.7)	0.02
Direct-acting antivirals before baseline, n (%)	108 (10.5)	1627 (10.4)	0.00	95 (9.8)	1606 (10.5)	0.02

*The variables for individual antidiabetic medications are not balanced after weighting because the variable for antidiabetic medications other than metformin or GLP1 agonists was modelled as a binary exposure.
.ALD, alcohol-related liver disease; GLP1, glucagon-like peptide-1 receptor; HCC, hepatocellular carcinoma; MASLD, metabolic dysfunction-associated steatotic liver disease; SMD, standardised mean difference.

MALO with 2 years of follow-up would need enrolment of at least 10 776 patients per arm to provide evidence for such effect.

DISCUSSION

We emulated a nationwide target trial to answer the question whether GLP1 agonists can prevent development of MALO in patients with chronic liver diseases and type 2 diabetes. The main finding was that the 10-year risk of MALO was 49% lower in patients who initiated and adhered to GLP1 agonists compared with non-initiators in the per-protocol analysis, but the estimates from the intention-to-treat analysis were imprecise with a 95% CI for the RR ranging from 0.50 to 1.32.

A placebo-controlled randomised phase II trial of 320 patients with MASH reported that the proportion achieving MASH resolution without worsening fibrosis after 72 weeks was more than tripled in the arm receiving 0.4 mg of the GLP1 agonist semaglutide compared with placebo (59% vs 17%).[14] We followed patients with any chronic liver disease and type 2 diabetes for up to 10 years and found a similarly strong effect of GLP1 agonists on the risk of MALO in per-protocol analysis, but an imprecise intention-to-treat effect. However, MALO risk was similar in the initiator arm when restricting the analysis to patients with MASLD, although this should be interpreted cautiously as we noted wide CIs for these estimates. The phase II trial only included patients with fibrosis stages F2–F3 in the primary outcome analysis, whereas we included patients of any fibrosis stage including compensated cirrhosis.[14] Our study did not provide support for a protective effect of GLP1 agonists in the subgroup with compensated cirrhosis, which is in line with a recent

Table 3 Risk of major adverse liver outcomes according to intention-to-treat and per-protocol analyses

	No of events, GLP1 agonist initiators	No of events, non-initiators	Absolute risk (%), GLP1 agonist initiators (95% CI)	Absolute risk (%), non-initiators (95% CI)	Risk difference (%) (95% CI)	Risk ratio (95% CI)
Intention to treat						
2 years	16	436	2.3 (1.0 to 3.6)	3.2 (2.9 to 3.5)	−0.9 (−2.2 to 4.3)	0.72 (0.31 to 1.13)
4 years	31	756	6.6 (3.4 to 9.9)	6.4 (6.0 to 6.9)	0.2 (−3.1 to 3.5)	1.03 (0.52 to 1.54)
6 years	36	933	8.0 (4.5 to 11.6)	9.1 (8.4 to 9.7)	−1.0 (−4.6 to 2.5)	0.89 (0.50 to 1.27)
8 years	41	1036	11.6 (7.0 to 16.3)	11.8 (10.9 to 12.8)	−0.2 (−4.8 to 4.4)	0.98 (0.59 to 1.38)
10 years	42	1079	13.3 (7.4 to 19.2)	14.6 (13.1 to 16.1)	−1.3 (−7.2 to 4.6)	0.91 (0.50 to 1.32)
Per protocol						
2 years	13	436	2.6 (0.9 to 4.3)	3.2 (2.9 to 3.5)	−0.6 (−2.4 to 1.1)	0.80 (0.26 to 1.34)
4 years	19	756	4.5 (2.1 to 7.0)	6.3 (5.9 to 6.9)	−1.9 (−4.3 to 0.6)	0.71 (0.32 to 1.10)
6 years	21	933	5.4 (2.7 to 8.1)	9.0 (8.4 to 9.7)	−3.6 (−6.4 to 0.8)	0.60 (0.29 to 0.90)
8 years	22	1036	7.2 (2.5 to 11.9)	11.7 (10.8 to 12.7)	−4.6 (−9.3 to 0.1)	0.61 (0.21 to 1.01)
10 years	22	1079	7.4 (2.1 to12.6)	14.4 (12.9 to 15.9)	−7.1 (−12.5 to 1.6)	0.51 (0.14 to 0.88)

.GLP1, glucagon-like peptide-1 receptor.

phase II trial in patients with MASH-related compensated cirrhosis.[41]

Based on the observed probability of MALO, our data indicate that a clinical trial of GLP1 agonists in patients with any chronic liver disease and type 2 diabetes using 2-year MALO risk as the outcome would demand an immense number of patients, at least 10 776 patients per arm. This can be contrasted to the ongoing phase III trial of semaglutide looking at both histopathological endpoints (after 1.5 years) and MALO (after 4.5 years) that is planning to recruit 1200 patients with MASH (identifier NCT04822181). The emulation of a target trial is an appealing option to give timely answers to key research questions when data from large clinical trials are currently unavailable.[32] The difference between our intention-to-treat and per-protocol estimates is that many initiators stopped treatment without one of our prespecified contraindications. Patients who ended their treatment might have done so for good reasons, for example, severe gastrointestinal symptoms, but we were unable to capture this in our data. Moreover, since we lacked data to specifically determine whether GLP1 agonists were indicated for the non-initiators during follow-up, we assumed that it was indicated for all non-initiators that started treatment. Therefore, none of the non-initiators were censored for not adhering to protocol, whereas many initiators stopped treatment and were then censored in the per-protocol analysis. Large observational studies including detailed data with relevance for the choice of pharmacological treatment in type 2 diabetes, such as glycated haemoglobin, are warranted.

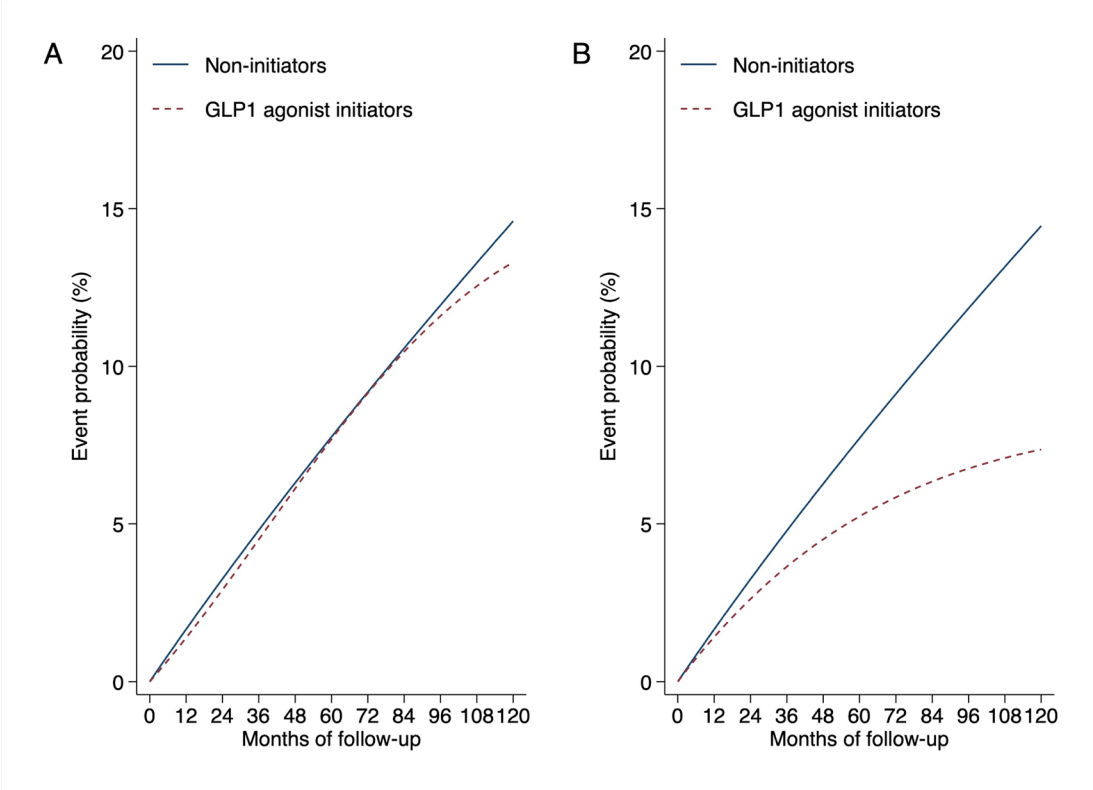

Figure 2 Inverse-probability weighted risk curves of major adverse liver outcomes comparing initiators of glucagon-like peptide-1 receptor (GLP1) agonists with non-initiators. (A) intention-to-treat effect, (B) per-protocol effect.

Table 4 Risk of major adverse liver outcomes at 10 years according to intention-to-treat analyses in subgroups

	No of events, GLP1 agonist initiators	No of events, non-initiators	Absolute risk (%), GLP1 agonist initiators (95% CI)	Absolute risk (%), non-initiators (95% CI)	Risk difference (%) (95% CI)	Risk ratio (95% CI)
Liver disease aetiology						
MASLD	21	303	15.8 (5.0 to 26.5)	11.2 (9.1 to 13.2)	4.6 (-5.5 to 14.8)	1.41 (0.53 to 2.30)
Other than MASLD	21	688	13.1 (2.6 to 23.6)	17.3 (14.5 to 20.2)	–4.2 (-14.9 to 6.4)	0.76 (0.15 to 1.36)
Cirrhosis status at baseline						
Compensated cirrhosis	23	286	36.5 (6.6 to 66.3)	34.6 (26.8 to 42.5)	1.9 (-27.6 to 31.4)	1.05 (0.20 to 1.91)
No cirrhosis	19	546	8.9 (2.9 to 14.9)	10.9 (9.1 to 12.6)	–1.9 (-7.8 to 3.9)	0.82 (0.29 to 1.36)

.GLP1, glucagon-like peptide-1 receptor; MASLD, metabolic dysfunction-associated steatotic.

The main strength of our study was the use of an emulated target trial design to overcome common biases in observational analyses, including immortal time bias.[34] In fact, bias in observational studies often arise predominantly due to poor study design, rather than confounding due to lack of randomisation, and estimates from carefully emulated target trials closely resemble those from randomised trials.[36] There were two reasons why we designed a target trial that compared GLP1 agonists to non-initiators rather than an active comparator, another common approach.[42–44] First, it would have asked a different research question, that is, whether MALO risk differs from the active comparator. An ideal active comparator has an identical indication as the drug of interest and no effect on the outcome, but the effect of other antidiabetic medications on MALO risk is mostly unknown. Second, the average treatment effect in the whole study population is not identifiable when comparing to an active comparator.[45] We additionally fitted an inverse-probability weighted marginal structural model to account for possible time-varying confounders associated with adherence when estimating the per-protocol effect. Whereas the intention-to-treat effect is the most used causal contrast in clinical trials, the per-protocol analysis indicates what the effect of a treatment strategy would be if adhered to, which is of great importance in the real-world when physicians and patients decide on an appropriate treatment strategy.[38] When there is feedback between time-varying confounders and treatments, then standard regression models (eg, Cox) will produce biased estimates.[39] The family of g methods has been developed for this purpose (including inverse-probability weighted marginal structural models and standardisation).[39] In addition, we used the validated population-based Swedish national healthcare registers to minimise selection bias and estimate the long-term effect of GLP1 agonists.[23–25 29–31]

Some limitations should be acknowledged. Despite a median follow-up of 5–6 years, few initiators experienced MALO, yielding estimates with low precision and preventing estimation of per-protocol effects in subgroups. The low number of events in initiators during late follow-up (only one event in the last 4 years and none in the last 2 years) could possibly explain part of the per-protocol effect, however, the RRs were stable across time from year 6 and forward. Additionally, we lacked data on some important covariates. First, we had no data on fibrosis stage, beyond classifying patients as cirrhotic (F4) or non-cirrhotic (F0–F3). Trials in MASLD are usually confined to patients with F2–F3, or cirrhosis, but we likely included some patients with F0–F1 where MALO is unlikely to occur. Second, we had no laboratory data, such as glycated haemoglobin, to assess diabetes severity and need for escalating to second-line treatment with GLP1 agonists. Diabetes duration, microvascular complications and other antidiabetic medications were, however, used as proxies for diabetes severity. The E-value suggested that an unmeasured confounder would need to increase the probability of both initiating a GLP1 agonist and experiencing MALO more than threefold to fully explain the estimated per-protocol effect.[40] This suggests that the observed RR in the per-protocol analysis might be explained by residual confounding (eg, fibrosis stage). For example, a study of patients with biopsy-proven MASLD found that patients with F3 had a fourfold higher hazard of MALO than patients with F0.[17] However, to fully explain the estimated effect, an unmeasured confounder would also need to be three times more likely to occur in either group. The presence of compensated cirrhosis was balanced between groups and other parameters associated with fibrosis such as age and cardiovascular disease were also well balanced and thus suggests that large differences in fibrosis are unlikely. Additionally, we sampled 5% of non-initiators to increase computational efficiency. Point estimates were, however, similar when including all non-initiators.

In conclusion, the risk of MALO in patients with chronic liver diseases and type 2 diabetes was lower if they initiated a GLP1 agonist and adhered to this treatment over time. The data were, however, not compatible with a protective intention-to-treat effect. Randomised trials using MALO as an outcome might be unfeasible, motivating further large observational studies using appropriate methodology to further delineate the effect of GLP1 agonists on the risk of MALO, complementing future phase III trials of GLP1 agonists.

Correction notice This article has been corrected since it published Online First. The senior author statement has been added.

Twitter Axel Wester @wester_axel, Ying Shang @YingShang1 and Hannes Hagström @hanneshagstrom

Acknowledgements AW was supported by the Syskonen Svensson foundation for medical research (2021-00284), Mag-tarmfonden, the Bengt Ihre foundation (SLS-973809), Professor Nanna Svartz foundation (2022-00448) and the Stockholm County Council (FoUI-985859). AAM was supported by the Strategic Research Program in Epidemiology at Karolinska Institutet, Forte (2020-00029), and the Swedish Research Council (2021-02236). HH was supported by the Stockholm County Council (FoUI-960537), the Swedish Research Council (2021-01293) and the Swedish Cancer Society (22210 Pj01H).

Contributors Contributor and guarantor information: Study conception and design: All Acquisition of data: HH Statistical analysis: AW and AAM. Analysis and interpretation of data: All Drafting of manuscript: AW. Critical revision: AllGuarantors of article: AW and HH.

Funding AW was supported by the Syskonen Svensson foundation for medical research (2021-00284), Mag-tarmfonden, the Bengt Ihre foundation (SLS-973809), Professor Nanna Svartz foundation (2022-00448), and the Stockholm County Council (FoUI-985859). AAM was supported by the Strategic Research Program in Epidemiology at Karolinska Institutet, Forte (2020-00029), and the Swedish Research Council (2021-02236). HH was supported by the Stockholm County Council (FoUI-960537), the Swedish Research Council (2021-01293), and the Swedish Cancer Society (2 2210 Pj 01 H). The funders had no role in the conduct of this study.

Disclaimer The funders had no role in the conduct of this study.

Competing interests HH: institution has received research grants from Astra Zeneca, EchoSens, Gilead, Intercept, MSD, Novo Nordisk and Pfizer, all outside the current study. The other authors declare no conflicts.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and the Regional Ethics Board of Stockholm approved the study and waived informed consent (protocol number 2017/1019-31/1).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. No data are available due to Swedish regulations.

ORCID iDs
Axel Wester http://orcid.org/0000-0003-3634-6591
Ying Shang http://orcid.org/0000-0002-1496-1799
Hannes Hagström http://orcid.org/0000-0002-8474-1759

REFERENCES

1 Collaborators GBDC. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:245–66.
2 Riazi K, Azhari H, Charette JH, *et al*. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2022;7:851–61.
3 Younossi ZM, Golabi P, de Avila L, *et al*. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. *J Hepatol* 2019;71:793–801.
4 Younossi ZM, Koenig AB, Abdelatif D, *et al*. Global epidemiology of Nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
5 Estes C, Anstee QM, Arias-Loste MT, *et al*. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United kingdom, and United States for the period 2016-2030. *J Hepatol* 2018;69:896–904.
6 Ajmera V, Cepin S, Tesfai K, *et al*. A prospective study on the prevalence of NAFLD, advanced fibrosis, cirrhosis and hepatocellular carcinoma in people with type 2 diabetes. *J Hepatol* 2023;78:471–8.
7 Björkström K, Franzén S, Eliasson B, *et al*. Risk factors for severe liver disease in patients with type 2 diabetes. *Clin Gastroenterol Hepatol* 2019;17:2769–75.
8 Elkrief L, Rautou PE, Sarin S, *et al*. Diabetes mellitus in patients with cirrhosis: clinical implications and management. *Liver Int* 2016;36:936–48.
9 Rao Kondapally Seshasai S, Kaptoge S, Thompson A, *et al*. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* 2011;364:829–41.
10 Wilding JPH, Batterham RL, Calanna S, *et al*. Once-weekly Semaglutide in adults with overweight or obesity. *N Engl J Med* 2021;384:989–1002.
11 Andreadis P, Karagiannis T, Malandris K, *et al*. Semaglutide for type 2 diabetes mellitus: A systematic review and meta-analysis. *Diabetes Obes Metab* 2018;20:2255–63.
12 Konerman MA, Jones JC, Harrison SA. Pharmacotherapy for NASH: current and emerging. *J Hepatol* 2018;68:362–75.
13 Armstrong MJ, Gaunt P, Aithal GP, *et al*. Liraglutide safety and efficacy in patients with non-alcoholic Steatohepatitis (LEAN): a Multicentre, double-blind, randomised, placebo-controlled phase 2 study. *Lancet* 2016;387:679–90.
14 Newsome PN, Buchholtz K, Cusi K, *et al*. A placebo-controlled trial of subcutaneous Semaglutide in Nonalcoholic Steatohepatitis. *N Engl J Med* 2021;384:1113–24.
15 Dufour J-F, Anstee QM, Bugianesi E, *et al*. Current therapies and new developments in NASH. *Gut* 2022;71:2123–34.
16 Dulai PS, Singh S, Patel J, *et al*. Increased risk of mortality by fibrosis stage in Nonalcoholic fatty liver disease: systematic review and meta-analysis. *Hepatology* 2017;65:1557–65.
17 Hagström H, Nasr P, Ekstedt M, *et al*. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J Hepatol* 2017;67:1265–73.
18 Åberg F, Byrne CD, Pirola CJ, *et al*. Alcohol consumption and metabolic syndrome: clinical and Epidemiological impact on liver disease. *J Hepatol* 2023;78:191–206.
19 Hagström H, Nasr P, Ekstedt M, *et al*. Alcohol consumption in concomitant liver disease: how much is too much. *Curr Hepatol Rep* 2017;16:152–7.
20 Ramesh S, Sanyal AJ. Hepatitis C and Nonalcoholic fatty liver disease. *Semin Liver Dis* 2004;24:399–413.
21 Israelsen M, Juel HB, Detlefsen S, *et al*. Metabolic and genetic risk factors are the strongest predictors of severity of alcohol-related liver fibrosis. *Clin Gastroenterol Hepatol* 2022;20:1784–94.
22 Hagström H, Shang Y, Wester A, *et al*. Cohort profile: decoding the epidemiology of liver disease in Sweden (DELIVER). *Scand J Gastroenterol* 2022;57:978–83.
23 Barlow L, Westergren K, Holmberg L, *et al*. The completeness of the Swedish cancer register: a sample survey for year 1998. *Acta Oncol* 2009;48:27–33.
24 Ludvigsson JF, Almqvist C, Bonamy A-KE, *et al*. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;31:125–36.
25 Ludvigsson JF, Andersson E, Ekbom A, *et al*. External review and validation of the Swedish National inpatient register. *BMC Public Health* 2011;11:450.
26 Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, *et al*. The Swedish personal identity number: possibilities and pitfalls in Healthcare and medical research. *Eur J Epidemiol* 2009;24:659–67.
27 Ludvigsson JF, Svedberg P, Olén O, *et al*. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *Eur J Epidemiol* 2019;34:423–37.
28 Wettermark B, Hammar N, Fored CM, *et al*. The new Swedish prescribed drug register--opportunities for Pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;16:726–35.
29 Brooke HL, Talbäck M, Hörnblad J, *et al*. The Swedish cause of death register. *Eur J Epidemiol* 2017;32:765–73.
30 Åström H, Wester A, Hagström H. Administrative coding for non-alcoholic fatty liver disease is accurate in Swedish patients. *Scand J Gastroenterol* 2023;58:931–6.
31 Bengtsson B, Askling J, Ludvigsson JF, *et al*. Validity of administrative codes associated with cirrhosis in Sweden. *Scand J Gastroenterol* 2020;55:1205–10.
32 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
33 Matthews AA, Danaei G, Islam N, *et al*. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ* 2022;378:e071108.
34 Hernán MA, Sauer BC, Hernández-Díaz S, *et al*. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016;79:70–5.
35 García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol* 2017;32:495–500.
36 Dickerman BA, García-Albéniz X, Logan RW, *et al*. Avoidable flaws in observational analyses: an application to Statins and cancer. *Nat Med* 2019;25:1601–6.
37 Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics - Simulation and Computation* 2009;38:1228–34.
38 Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med* 2017;377:1391–8.
39 Naimi AI, Cole SR, Kennedy EH. An introduction to G methods. *Int J Epidemiol* 2017;46:756–62.
40 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med* 2017;167:268–74.
41 Loomba R, Abdelmalek MF, Armstrong MJ, *et al*. Semaglutide 2.4 mg once weekly in patients with non-alcoholic Steatohepatitis-related cirrhosis: a randomised, placebo-controlled phase 2 trial. *Lancet Gastroenterol Hepatol* 2023;8:511–22.
42 Abrahami D, McDonald EG, Schnitzer ME, *et al*. Proton pump inhibitors and risk of gastric cancer: population-based cohort study. *Gut* 2022;71:16–24.
43 Abrahami D, McDonald EG, Schnitzer ME, *et al*. Proton pump inhibitors and risk of colorectal cancer. *Gut* 2022;71:111–8.
44 Suissa S, Suissa A. Proton-pump inhibitors and increased gastric cancer risk: time-related biases. *Gut* 2018;67:2228–9.
45 Huitfeldt A, Hernan MA, Kalager M, *et al*. Comparative effectiveness research using observational data: active Comparators to emulate target trials with inactive Comparators. *EGEMS (Wash DC)* 2016;4:1234.