

Macrovascular and microvascular outcomes of metabolic surgery versus GLP-1 receptor agonists in patients with diabetes and obesity

Received: 25 October 2024

A list of authors and their affiliations appears at the end of the paper

Accepted: 15 July 2025

Published online: 16 September 2025

 Check for updates

Both metabolic surgery and glucagon-like peptide-1 (GLP-1) receptor agonists (RAs) improve cardiometabolic outcomes, but their long-term outcomes have not been directly compared. Here, we compared macrovascular and microvascular outcomes in 1,657 patients (65.7% female) with type 2 diabetes and obesity who underwent metabolic surgery with 2,275 similar patients (53.5% female) who received treatment with GLP-1 RAs. Using a doubly robust estimation method to balance baseline characteristics between groups, we examined the time to all-cause mortality, incident major adverse cardiovascular events (MACE), nephropathy and retinopathy over a median follow-up of 5.9 years. The 10-year cumulative incidence of all-cause mortality was 9.0% (95% confidence interval (CI) 6.8–10.8%) in the metabolic surgery group and 12.4% (95% CI 9.9–15.2%) in the GLP-1 RA group (adjusted hazard ratio (HR) 0.68 (95% CI 0.48–0.96), $P = 0.028$). Compared with the GLP-1 RA group, metabolic surgery was also associated with a lower risk of MACE (adjusted HR 0.65; 95% CI 0.51–0.82; $P < 0.001$), nephropathy (adjusted HR 0.53; 95% CI 0.43–0.67; $P < 0.001$) and retinopathy (adjusted HR 0.46; 95% CI 0.29–0.75; $P = 0.002$). These findings indicate that even with the availability of GLP-1 RAs, metabolic surgery remains superior to medical treatment. Future studies should compare the cardiometabolic outcomes of metabolic surgery with newer GLP-1 RAs that are more effective for weight reduction.

Cardiovascular disease, nephropathy and retinopathy are major causes of morbidity and premature mortality among patients with obesity and diabetes. Small randomized clinical trials (RCTs) have documented the significant effect of metabolic surgery on excess body weight and diabetes control^{1–3}. Larger-scale observational studies further showed that metabolic surgery, compared with nonsurgical management, is associated with a significantly lower risk of incident macrovascular and microvascular disease^{4–7}.

Over the past decade, glucagon-like peptide-1 (GLP-1) receptor agonists (RAs) have emerged as a recommended pharmacological option for glycemic control and cardioprotective therapy in patients

with type 2 diabetes mellitus (T2DM) and obesity⁸. This has been backed by evidence from RCTs showing that GLP-1 RAs prevent cardiovascular events, particularly in those with established atherosclerotic cardiovascular disease^{9–12}.

While both metabolic surgery and GLP-1 RAs improve cardiometabolic outcomes, their impact on incident macrovascular and microvascular disease, including major adverse cardiovascular events (MACE), nephropathy, retinopathy and mortality has not been directly compared. It is challenging and costly to conduct sufficiently powered RCTs with adequate follow-up in this setting.

Table 1 | Baseline characteristics of patients in the metabolic surgery and GLP-1 RA groups at the index date

Baseline variable	Unweighted			Weighted ^a		
	Surgical (n=1,657)	GLP-1 RA (n=2,275)	Standardized difference ^b	Surgical (n=1,657)	GLP-1 RA (n=2,275)	Standardized difference ^b
Demographics	Age (years), mean (s.d.)	51 (11.3)	56.6 (10.2)	0.528	53.7 (10.6)	53.7 (10.7)
	Annual zip code income (thousands of US\$), mean (s.d.)	53.7 (19.6)	56.2 (19.5)	0.127	54.9 (19.8)	54.9 (19.5)
	(Missing)	45 (2.7)	10 (0.4)	0	0	0
	Index date, mean (s.d.)	2014 (2.2)	2016.3 (1.3)	1.247	2014.1 (2.2)	2016.2 (1.3)
	Location	Florida	429 (25.9)	282 (12.4)	0.348	(18.4)
		Ohio	1,228 (74.1)	1,993 (87.6)		(81.6)
	Race ^c	Black	332 (20)	317 (13.9)	0.168	(16.8)
		Missing	30 (1.8)	56 (2.5)		(2.1)
		Other	53 (3.2)	68 (3)		(2.9)
		White	1,242 (75)	1,834 (80.6)		(78.1)
	Sex	Female	1,089 (65.7)	1,218 (53.5)	0.250	(63.2)
		Male	568 (34.3)	1,057 (46.5)		(36.8)
	Smoking status	Current	41 (2.5)	206 (9.1)	0.379	(4.3)
		Former	646 (39)	876 (38.5)		(39.9)
		Missing	83 (5)	17 (0.7)		(1.6)
		Never	887 (53.5)	1,176 (51.7)		(54.1)
Clinical and laboratory data	BMI (kg m^{-2}), mean (s.d.)	46.3 (8.8)	38.6 (6.9)	-0.975	42.4 (6.2)	42.4 (8.5)
	Systolic BP (mmHg), mean (s.d.)	130.9 (15.8)	128.9 (15.5)	-0.129	129.6 (15.7)	129.6 (15.2)
	(Missing)	0	12 (0.5)	0	0	0
	HbA1c (%), mean (s.d.)	7.5 (1.6)	7.6 (1.5)	0.082	7.6 (1.5)	7.6 (1.6)
	(Missing)	161 (9.7)	336 (14.8)	0	0	0
	HDL-C (mg dl^{-1}), mean (s.d.)	44.5 (11.6)	42.8 (12.3)	-0.142	44 (11.5)	44 (12.6)
	(Missing)	183 (11)	464 (20.4)	0	0	0
	LDL-C (mg dl^{-1}), mean (s.d.)	95.8 (33.5)	85.6 (33)	-0.308	91.5 (32.7)	91.5 (34.5)
	(Missing)	195 (11.8)	464 (20.4)	0	0	0
	Triglycerides (mg dl^{-1}), mean (s.d.)	176.7 (141.4)	177.2 (155.5)	0.004	178 (120.2)	178 (167.4)
	(Missing)	183 (11)	466 (20.5)	0	0	0
	UACR (mg g^{-1}), mean (s.d.)	98.5 (362.7)	87.6 (388.9)	-0.029	63.1 (249.6)	63.1 (274.7)
	(Missing)	1,040 (62.8)	1,327 (58.3)	0	0	0
	eGFR ^d ($\text{mL min}1.73 \text{m}^{-2}$), mean (s.d.)	90.9 (22.8)	88.6 (21.4)	-0.103	89.8 (21.5)	89.8 (22.3)
	(Missing)	0	139 (6.1)	0	0	0
Medication history	Insulin	444 (26.8)	758 (33.3)	0.143	(30.8)	(30.8)
	Count of non-insulin diabetes medications (numerical), mean (s.d.)	0.9 (0.8)	1.2 (0.9)	0.345	1.1 (0.9)	1.1 (0.9)
	Count of non-insulin diabetes medications	0	540 (32.6)	0.367	(26.9)	(27.6)
		1	780 (47.1)	969 (42.6)		(47.3)
		2	265 (16)	619 (27.2)		(19.5)
		3	70 (4.2)	153 (6.7)		(5.9)
		4	2 (0.1)	27 (1.2)		(0.3)
	Lipid-lowering drugs	816 (49.2)	1,502 (66)	0.345	(58.5)	(58.5)
	RAAS Inhibitors ^e	894 (54)	1,486 (65.3)	0.233	(59.7)	(59.7)
	Other antihypertensive drugs	878 (53)	1,151 (50.6)	-0.048	(52.2)	(52.2)
Aspirin		366 (22.1)	566 (24.9)	0.066	(24.5)	(24.5)
	Warfarin	57 (3.4)	53 (2.3)	-0.066	(2.7)	(2.7)

Table 1 (continued) | Baseline characteristics of patients in the metabolic surgery and GLP-1 RA groups at the index date

Baseline variable	Unweighted			Weighted ^a		
	Surgical (n=1,657)	GLP-1 RA (n=2,275)	Standardized difference ^b	Surgical (n=1,657)	GLP-1 RA (n=2,275)	Standardized difference ^b
Medical history	Atrial fibrillation	75 (4.5)	128 (5.6)	0.05	(4.7)	(4.7)
	Cerebrovascular disease	26 (1.6)	39 (1.7)	0.011	(1.4)	(1.4)
	Chronic obstructive pulmonary disease	115 (6.9)	151 (6.6)	-0.012	(6.5)	(6.5)
	Coronary arterial disease	121 (7.3)	224 (9.8)	0.091	(8.5)	(8.5)
	Dyslipidemia	1,031 (62.2)	1,702 (74.8)	0.274	(70.4)	(70.4)
	Heart failure	93 (5.6)	90 (4)	-0.078	(4.7)	(4.7)
	Hypertension	1,238 (74.7)	1,728 (76)	0.029	(76.3)	(76.3)
	Myocardial infarction	26 (1.6)	62 (2.7)	0.08	(1.9)	(1.9)
	Nephropathy	243 (14.7)	323 (14.2)	-0.013	(14.6)	(14.6)
	Neuropathy	153 (9.2)	377 (16.6)	0.22	(13.4)	(13.4)
Physical examination	Peripheral arterial disease	52 (3.1)	105 (4.6)	0.077	(4.0)	(4.0)
	Retinopathy	35 (2.1)	142 (6.2)	0.208	(3.2)	(3.2)

Data, including missing data, are *n* (%) unless stated otherwise. ^aBased on the average weighted summary values over the imputation datasets. After overlap weighting, a single individual no longer represents a single data entity; thus, raw counts are not reported after overlap weighting. ^bAbsolute value of the between-group difference in means or proportions (metabolic surgery–GLP-1 RA group) divided by the pooled s.d. ^cRace was obtained from electronic health records and based on patients selecting from fixed categories. ^dCalculated using the 2021 CKD-EPI creatinine equation²³. ^eRAAS inhibitors include angiotensin-converting enzyme inhibitors and angiotensin receptor blockers. Abbreviations: HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; UACR, urine albumin-to-creatinine ratio.

The M6 (Macrovascular and Microvascular Morbidity and Mortality after Metabolic surgery versus Medicines) observational study aimed to investigate the association between metabolic surgery and long-term risk of all-cause mortality, MACE, nephropathy and retinopathy in patients with obesity and T2DM, compared with precisely balanced patients who were medically managed with GLP-1 RAs. The study also examined the association of metabolic surgery with measures of diabetes control and medication use. The study tried to examine whether there is any benefit from metabolic surgery in the age of GLP-1 RA medications.

Results

Participants

A total of 3,932 adult patients (2,307 (58.7%) female; 3,076 (78.2%) White, 649 (16.5%) Black, 121 (3.1%) other races; mean age, 54.2 years (s.d. 11.0); mean body mass index (BMI), 41.8 kg m⁻² (s.d. 8.6)) with obesity and T2DM who received care at the Cleveland Clinic Health System (CCHS) in the United States from 2010 to 2017 were studied (Extended Data Fig. 1).

There were 1,657 patients in the metabolic surgery group (including 1,015 (61.3%) Roux-en-Y gastric bypass (RYGB) and 642 (38.7%) sleeve gastrectomy) and 2,275 patients in the GLP-1RA group. The GLP-1 RA medications in the nonsurgical group during the follow-up were liraglutide in 1,488 (65.4%), dulaglutide in 1,106 (48.6%), exenatide in 740 (32.5%), semaglutide in 602 (26.5%), tirzepatide in 101 (4.4%) and lixisenatide in 65 (2.9%) patients. In the metabolic surgery group, 6.0% of patients filled a prescription for GLP-1RAs in the 6 months leading to the surgery; that number decreased to 2.4% at 6 months after the surgery and continued to remain low (1.5%) at 1-year after surgery (Supplementary Fig. 1).

The distribution of 33 a priori-identified baseline covariates for the metabolic surgery and the GLP-1 RA group was precisely balanced (standardized differences = 0) after overlap weighting (Table 1). Supplementary Table 1 details the baseline socioeconomic characteristics of the surgical and nonsurgical groups.

The median follow-up for the entire cohort was 5.9 years (interquartile range (IQR) 4.4–7.6 years), including 6.1 years (IQR 2.3–8.7 years) for the metabolic surgery group and 5.9 years (IQR 5.0–7.0 years)

for the GLP-1 RA group. Surveillance in both groups followed a similar pattern before and after the index dates (Supplementary Fig. 2).

All-cause mortality

At the end of the study follow-up period (31 December 2022), 108 individuals in the metabolic surgery group and 153 in the GLP-1 RA group died. In the overlap weighted analysis, the cumulative incidence of all-cause mortality at 10 years was 9.0% (95% CI 6.8–10.8%) in the metabolic surgery group and 12.4% (95% CI 9.9–15.2%) in the GLP-1RA group. In the doubly robust analysis, the adjusted absolute risk difference was 3.8% (95% CI 0.3–7.2%); adjusted HR 0.68 (95% CI 0.48–0.96); *P* = 0.028 (Fig. 1a and Table 2). Separation of Kaplan–Meier curves favoring metabolic surgery was observed 4 years after the index date.

MACE

During the study follow-up, in the unweighted dataset, 220 patients in the metabolic surgery group and 488 patients in the GLP-1 RA group experienced incident MACE, which was a composite of four adverse events (coronary artery events, cerebrovascular events, heart failure or atrial fibrillation) (Extended Data Table 1).

In the overlap weighted analysis, the cumulative incidence of MACE at 10 years was 23.7% (95% CI 20.0–27.6%) in the metabolic surgery group and 34.0% (95% CI 28.1–44.2%) in the GLP-1 RA group. In the doubly robust analysis, the adjusted absolute risk difference was 8.8% (95% CI 3.8–13.3%); adjusted HR 0.65 (95% CI 0.51–0.82); *P* < 0.001 (Fig. 1b and Table 2).

The 10-year cumulative incidence estimates (Kaplan–Meier) of the MACE components are presented in the Extended Data Table 2.

Nephropathy

By the end of the M6 study, 218 patients in the metabolic surgery group and 440 patients in the GLP-1RA group experienced incident nephropathy, defined as a decline of estimated glomerular filtration rate (eGFR) of 40% or more or to <15 ml min⁻¹ 1.73 m⁻², initiation of dialysis or kidney transplant (Extended Data Table 3).

In the overlap weighted analysis, the cumulative incidence of nephropathy at 10 years was 21.4% (95% CI 17.8–24.7%) in the metabolic surgery group and 37.0% (95% CI 30.2–45.1%) in the GLP-1RA group. In

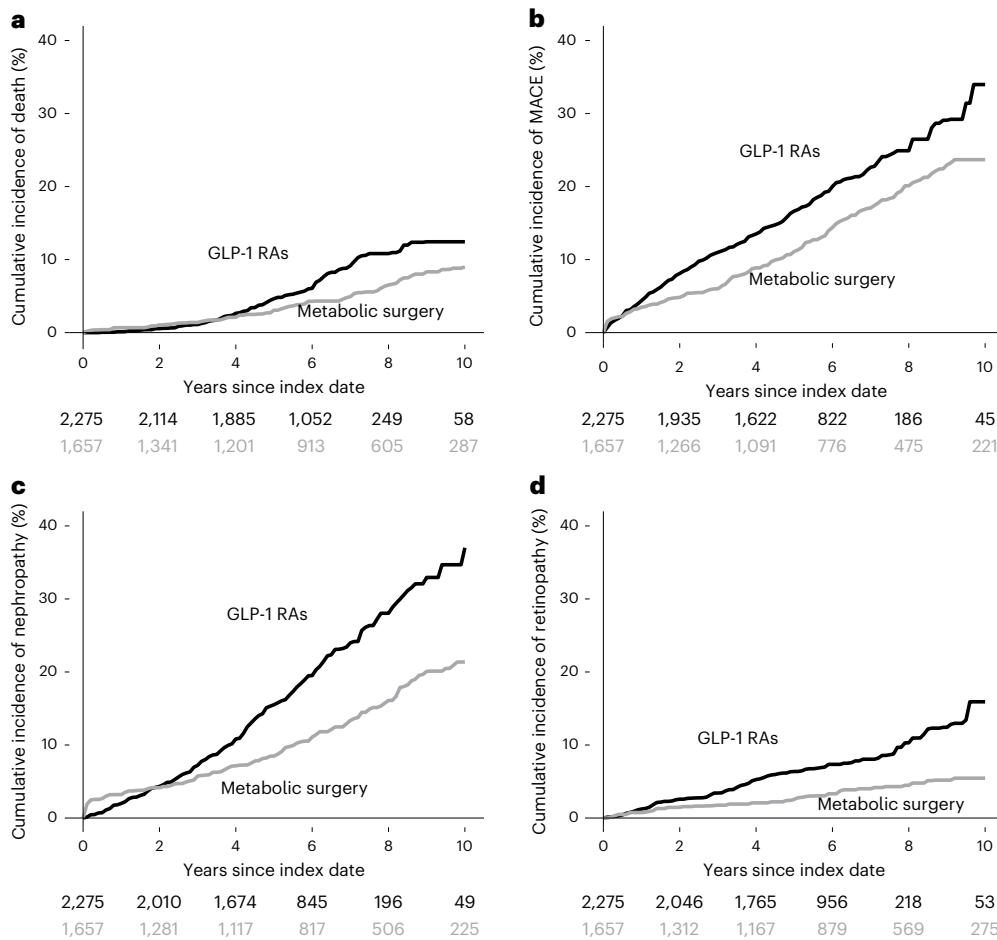


Fig. 1 | Ten-year cumulative incidence estimates for primary and secondary end points. a–d, Kaplan–Meier curves showing the cumulative incidence of all-cause mortality (a), incident MACE (b), incident nephropathy (c) and incident retinopathy (d) in the metabolic surgery and the GLP-1RA groups.

Table 2 | Incidence estimates, absolute risk differences and hazard ratios (HRs) for macrovascular and microvascular outcomes for metabolic surgery patients versus nonsurgical patients

Outcome	Metabolic surgery (n=1,657)			GLP-1 RA (n=2,275)			Unadjusted absolute 10-year risk difference (%) with 95% CI ^b	Adjusted absolute 10-year risk difference (%) with 95% CI ^c	Adjusted HR (95% CI) ^d	P ^d
	n with event	Rate per 100 patient-years ^a	Cumulative incidence (%) with 95% CI at 10 years ^b	n with event	Rate per 100 patient-years ^a	Cumulative incidence (%) with 95% CI at 10 years ^b				
All-cause mortality	108	0.95	9.0 (6.8–10.8)	153	1.17	12.4 (9.9–15.2)	3.5 (0.6–7.4)	3.8 (0.3–7.2)	0.68 (0.48–0.96)	0.028
MACE ^e	220	2.68	23.7 (20.0–27.6)	488	3.73	34.0 (28.1–44.2)	10.3 (3.2–21.6)	8.8 (3.8–13.3)	0.65 (0.51–0.82)	<0.001
Nephropathy	218	2.33	21.4 (17.8–24.7)	440	3.68	37.0 (30.2–45.1)	15.6 (8.0–23.4)	12.6 (8.1–16.7)	0.53 (0.43–0.67)	<0.001
Retinopathy	52	0.58	5.5 (4.1–7.4)	191	1.30	15.9 (11.2–25.0)	10.5 (4.9–19.9)	6.4 (2.7–10.6)	0.46 (0.29–0.75)	0.002

^aBased on overlap weighted data by taking the total number of (weighted) events observed divided by the total (weighted) event time (for those with events) or follow-up time (for those censored). ^bBased on overlap weighted data. ^cEstimated using fully adjusted Cox models. ^dHRs (95% CIs) and P values from fully adjusted Cox models comparing the relative instantaneous risk of each outcome for surgical versus nonsurgical patients; the CIs and two-sided P values were computed with a normal approximation. ^eA composite end point that was defined as the first occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation. Exact P value for MACE = 3.6×10^{-6} ; exact P value for nephropathy = 6.4×10^{-6} .

the doubly robust analysis, the adjusted absolute risk difference was 12.6% (95% CI 8.1–16.7%); adjusted HR 0.53 (95% CI 0.43–0.67); $P < 0.001$ (Fig. 1c and Table 2).

Retinopathy

During the study follow-up, 52 patients in the metabolic surgery group and 191 patients in the GLP-1RA group developed retinopathy, defined as development of de novo retinopathy or progression of preexisting retinopathy (Extended Data Table 3).

In the overlap weighted analysis, the cumulative incidence of retinopathy at 10 years was 5.5% (95% CI 4.1–7.4%) in the metabolic

surgery group and 15.9% (95% CI 11.2–25.0%) in the GLP-1RA group. In the doubly robust analysis, the adjusted absolute risk difference was 6.4% (95% CI 2.7–10.6%); adjusted HR 0.46 (95% CI 0.29–0.75); $P = 0.002$ (Fig. 1d and Table 2).

Subgroup analysis

Interaction testing did not show statistically significant heterogeneity in the association of metabolic surgery with risk of all-cause mortality, MACE, nephropathy and retinopathy based on sex, age, race, smoking status, BMI, glycated hemoglobin (HbA1c) level, eGFR or use of insulin or renin–angiotensin–aldosterone system (RAAS) inhibitors (Fig. 2).

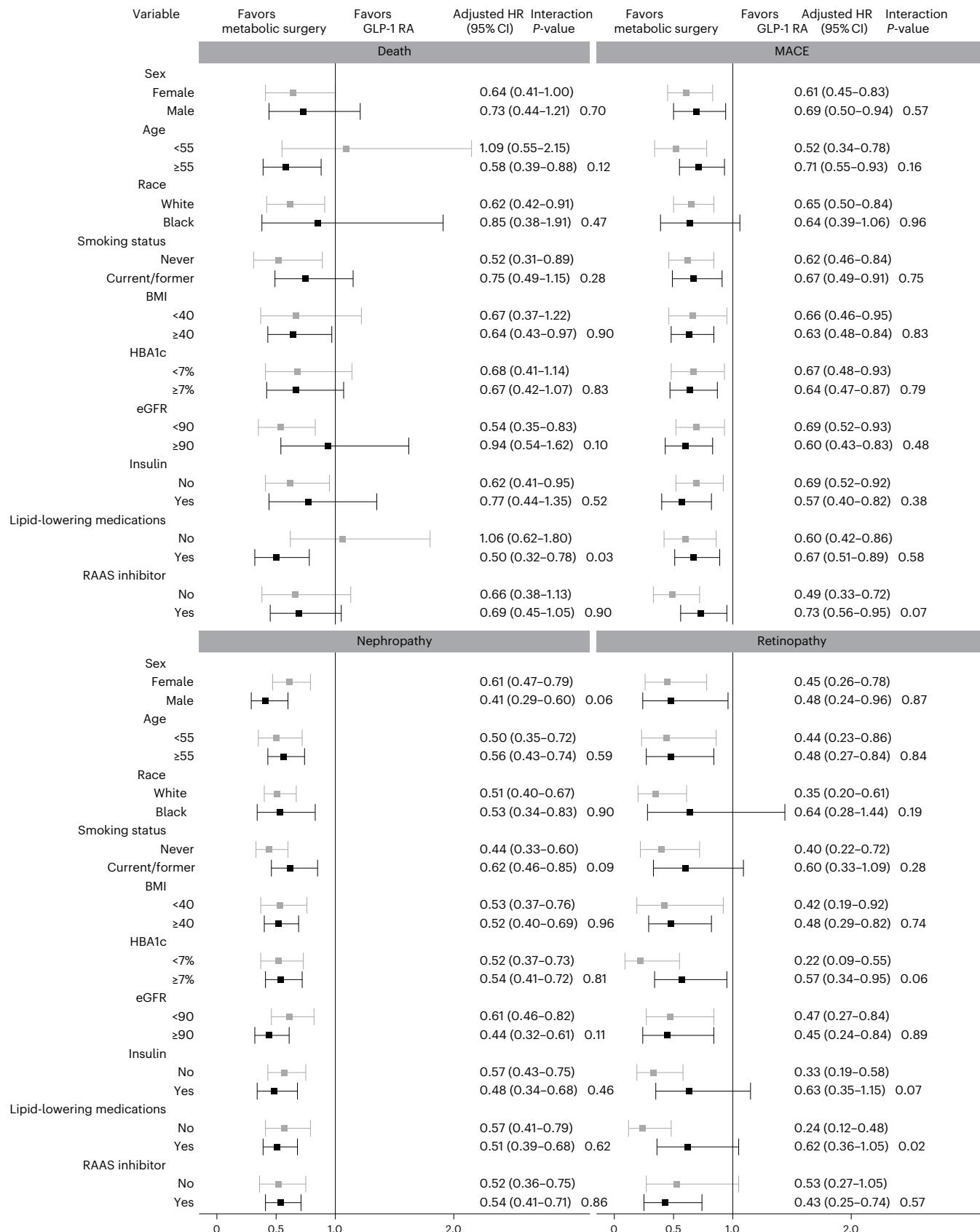


Fig. 2 | Association of metabolic surgery versus GLP-1RAs with primary and secondary end points in key subgroups in the fully adjusted Cox models. Adjusted HRs and 95% CIs are shown. Adjusted HRs were calculated after individually removing variables of interest from the fully adjusted Cox model and replacing it with the dichotomous subgroup variable and its interaction term with the treatment variable. For example, the continuous age covariate was replaced by a dichotomous version and its interaction term with the treatment. P-values for the interaction term between each variable and the surgical indicator from the fully adjusted model are also presented. The subgroup interactions were not adjusted for multiplicity and should be viewed cautiously.

term with the treatment variable. For example, the continuous age covariate was replaced by a dichotomous version and its interaction term with the treatment. P-values for the interaction term between each variable and the surgical indicator from the fully adjusted model are also presented. The subgroup interactions were not adjusted for multiplicity and should be viewed cautiously.

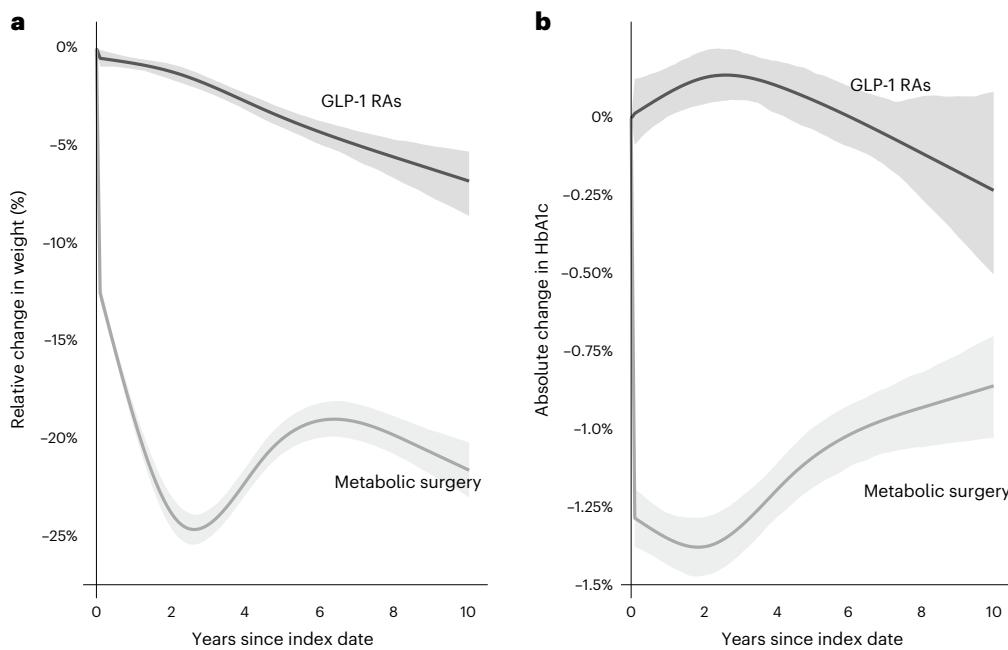


Fig. 3 | Mean trend curves of weight loss and HbA1c values over 10 years of follow-up. a,b, Plots displaying smoothed mean trends of percentage weight loss (a) and absolute HbA1c values (b) from baseline to follow-up. The shaded areas indicate the 95% CIs. The mean difference in total weight loss and HbA1c changes at 10 years between groups were estimated from a flexible, random-intercept

linear model, with four-knot restricted cubic spline for time \times treatment interaction on measurements. The 95% CIs were obtained by bootstrapping (100 bootstrap resampling) at the patient level and then using the percentile method across all imputation datasets.

The *P* value for the interaction was less than 0.05 for the association of metabolic surgery with all-cause mortality and retinopathy based on the use of lipid-lowering medications; the subgroup interactions were not adjusted for multiplicity and should be viewed cautiously.

The main results were stratified according to the surgical procedure type (Extended Data Table 4). Both RYGB and sleeve gastrectomy provided comparable benefits for all-cause mortality, MACE and retinopathy. Gastric bypass was associated with a larger reduction of incident nephropathy than sleeve gastrectomy.

Status of obesity, diabetes, and medication use over time

The mean body weight reduction at 10 years was 21.6% (95% CI 20.2–23.0%) in the metabolic surgery group and 6.8% (95% CI 5.3–8.6%) in the GLP-1RA group; the mean between-group difference in the overlap weighted analysis was 14.8% (95% CI 12.6–16.8%); *P* < 0.001 (Fig. 3a).

The absolute change in HbA1c level at 10 years was -0.86% (95% CI -1.03% to -0.70%) in the metabolic surgery group and -0.23% (95% CI -0.50% to 0.09%) in the GLP-1 RA group; the mean between-group difference in the overlap weighted analysis was -0.63% (95% CI -0.97% to -0.33%); *P* < 0.001 (Fig. 3b).

Throughout the study, the proportion of patients with active prescriptions for diabetes (including sodium-glucose cotransporter 2 inhibitors), blood pressure (BP) (including RAAS inhibitors) and cholesterol control was significantly higher in the nonsurgical group compared to the surgical group (Fig. 4, Extended Data Fig. 2 and Extended Data Table 5).

Adverse events after metabolic surgery

In the 90 days after metabolic surgery, 118 (7.1%) patients had serious complications and eight (0.48%) patients died. Serious complications included pulmonary adverse events ($n = 43$, 2.6%), bleeding requiring transfusion ($n = 40$, 2.4%), sepsis ($n = 26$, 1.6%), cardiac events ($n = 24$, 1.4%), venous thromboembolism ($n = 22$, 1.3%), gastrointestinal leak ($n = 15$, 0.9%), bowel obstruction requiring surgery ($n = 6$, 0.4%) and renal failure requiring dialysis ($n = 1$, 0.06%).

Sensitivity analyses

Although the primary analysis plan for this study was intention-to-treat, to address noncompliance in the nonsurgical group, we performed two sensitivity analyses by censoring nonsurgical patients who discontinued GLP-1RA medications. The main findings of the study were supported by these additional analyses (Extended Data Table 6).

Moreover, to assess the robustness of the identified associations, we censored surgical patients who continued to receive GLP-1 RAs after metabolic surgery or initiated GLP-1 RAs during the follow-up (Extended Data Table 7).

The E-value was used to assess the potential impact of unmeasured confounding factors. The magnitude of the associations of the known risk factors with the study endpoints was smaller than the estimated E-values for all-cause mortality (2.32), MACE (2.03), nephropathy (2.50) and retinopathy (3.74), which makes it unlikely that unmeasured confounders could eliminate the favorable association between metabolic surgery and the study endpoints (Extended Data Table 8).

Finally, a comparison of patients' health status between the study index date and either the GLP-1 RA initiation date (in the nonsurgical group) or the referral date for metabolic surgery (in the surgical group) showed a similar pattern of increased frequency of diagnosed medical conditions and medication use in both groups (Supplementary Table 2). In the surgical group, unweighted BMI decreased from 47.9 kg m^{-2} at the time of referral to 46.3 kg m^{-2} at the time of surgery. In the non-surgical group, unweighted BMI decreased from 39.3 kg m^{-2} at GLP-1 RA initiation to 38.6 kg m^{-2} on the index date. Changes in clinical and laboratory measures followed a similar trend in both groups. Nonsurgical patients experienced a greater reduction in HbA1c levels before the index date, from 8.0% to 7.6%. In the surgical group, HbA1c levels remained stable at 7.5% from the referral date to the day of surgery (Supplementary Table 2).

Discussion

In this observational study with a long follow-up time, metabolic surgery was associated with a significantly lower risk of all-cause mortality,

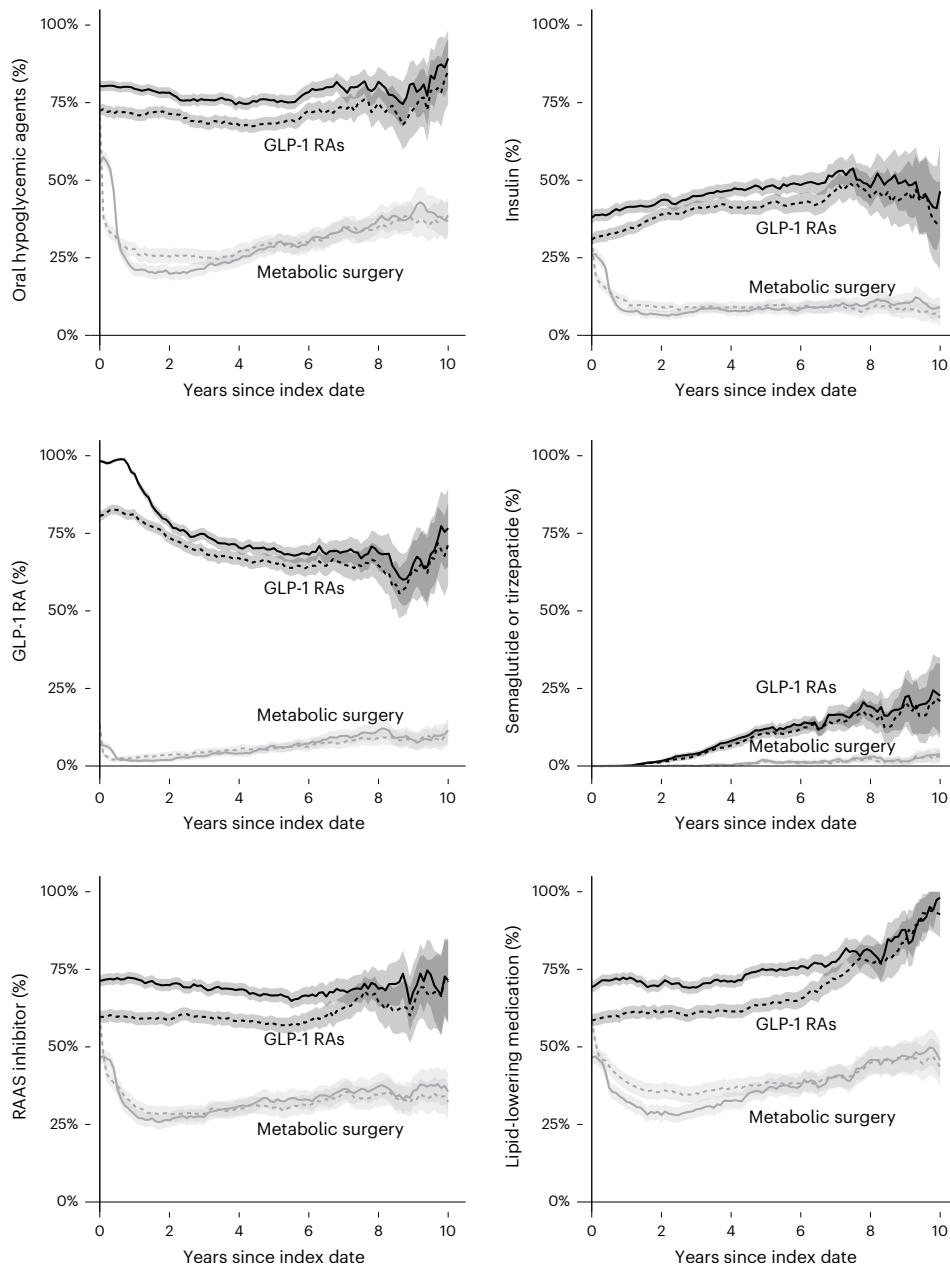


Fig. 4 | Proportions of patients who were prescribed or dispensed diabetes and cardiovascular drugs over 10 years of follow-up. Plots displaying the proportions of patients (with 95% point-wise CIs (shaded areas)) in the metabolic surgery and GLP-1 RA groups over time who were prescribed or dispensed with the indicated medications. For a given time point (for example, 1 year after the index date), prescription orders (dashed lines) were defined as the percentage of patients (of those who were followed up to at least that date) who had at least one prescription order for that medication, with start and end dates that

encompassed that time point (using the last follow-up date if the end date was missing). For a given time point (for example, 1 year after the index date), medications dispensed (solid lines) were defined as the percentage of patients (of those who were followed up to at least that date) who had at least one dispensing record for that medication in the preceding 6 months. Medication dispensing data were based on Surescripts data. RAAS inhibitors included angiotensin-converting enzyme inhibitors and angiotensin receptor blockers.

incident MACE, nephropathy and retinopathy compared to GLP-1 RA alone in a large and well-balanced sample of individuals with T2DM and obesity. Patients who received metabolic surgery, compared to the GLP-1 RA group, also experienced a larger reduction in total body weight, HbA1c levels and need of diabetes and cardiovascular medications from baseline to 10 years.

Studies comparing metabolic surgery and GLP-1RA medications have started to emerge. An observational propensity score-matched cohort study based on the Scandinavian national registries reported a 24% lower risk of MACE (HR 0.76 (95% CI 0.59–0.98)) in patients with T2DM and obesity who underwent metabolic surgery compared

with a nonsurgical group with T2DM from the general population treated with GLP-1 RAs (mainly liraglutide (81.7%); 0.8% semaglutide; no tirzepatide)¹³. Variables in that study were limited to those available in the registries and, for example, did not include weight and weight change in the nonsurgical group¹³. In another recent retrospective observational study based on data from a large health insurer in Israel, metabolic surgery was associated with a 57% greater reduction in primary incidence of heart failure (HR 0.43 (95% CI 0.27–0.68)) compared with treatment with GLP-1 RAs (mainly liraglutide (59.8%); 5.1% semaglutide; no tirzepatide) in patients with T2DM and obesity⁷.

Because sufficiently powered RCTs comparing metabolic surgery to GLP-1 RA medications with sufficient follow-up to document macrovascular and microvascular events are challenging and costly to conduct, carefully designed observational studies could provide reasonably strong evidence to inform treatment decisions. The current study included a larger group of nonsurgical patients who received semaglutide (26.5%) or tirzepatide (4.4%), investigated a wider range of macrovascular and microvascular outcomes, and used doubly robust estimation combining overlap weighting and outcome regression to minimize the effects of confounding factors when comparing outcomes in the metabolic surgery group and the GLP-1RA group^{14,15}. The first step in the doubly robust estimation—overlap weighting—attempts to mimic the key properties of RCTs, namely target population, covariate balance (i.e. standardized differences = 0 for the mean values of the 33 baseline covariates) and precision^{14,15}. Subsequently, outcome regression was used on 34 baseline covariates for further adjustment.

With the emergence of data showing that GLP-1 RAs induce clinically meaningful weight loss^{16,17} and prevent cardiovascular events^{9–12}, it has been suggested that they could potentially replace or obviate the need for metabolic surgery. If a patient has access to a GLP-1RA, is there any additional benefit to metabolic surgery? The present study can conclude that there are substantial additional benefits, including a survival benefit from metabolic surgery, in the era of GLP-1 RAs. Our sensitivity analyses showed the clinical benefits from metabolic surgery even in a subset of patients who were persistent with their GLP-1 RA medications. The lower rate of adverse clinical events after metabolic surgery may be related to substantial and sustained weight loss with subsequent improvement in metabolic, hemodynamic and neurohormonal pathways. A growing body of evidence over the last two decades suggests that anatomical changes in the gastrointestinal tract after metabolic surgery could be also responsible for metabolic and neurohormonal improvements, independent of weight loss^{18,19}. Such sustained improvements are more challenging to achieve with GLP-1 RA alone because non-persistence with them is common^{20,21}, which has been documented to result in weight regain and reversal of cardiometabolic improvements²². To be eligible for inclusion in this study, patients in the GLP-1RA group had to receive these medications before and after their index dates.

In addition to the risk of adverse cardiovascular events, the current study also examined the risk of microvascular outcomes of T2DM and obesity. Epidemiological studies support the negative impact of increased BMI on diabetic retinopathy²³. Increased BMI is also associated with an increased risk of diabetic and nondiabetic end-stage renal disease²⁴. An earlier observational study of adults with T2DM found that metabolic surgery was associated with a lower risk of nephropathy, retinopathy and neuropathy compared with usual care⁹. However, to our knowledge, the present study is the first to compare microvascular outcomes in patients with obesity and T2DM who underwent metabolic surgery versus those treated with GLP-1 RAs.

This study has several limitations. First, the most common GLP-1 RAs in the nonsurgical group were liraglutide (65.4%) and dulaglutide (48.6%), which have intermediate weight loss efficacy. Newer GLP-1 RA medications (for example, semaglutide and tirzepatide), which can provide more favorable weight loss and cardiometabolic outcomes^{12,17,25}, have become available to patients more recently and toward the end of the follow-up time (Fig. 4). Nevertheless, data from RCTs and observational studies showed smaller weight reduction in patients with T2DM, even with high dosages of newer GLP-1 RAs, such as semagludite^{26–28}. For example, the placebo-subtracted mean weight reduction in patients with overweight or obesity with semaglutide, 2.4 mg once a week, was 12.5% in those without T2DM (STEP 1 study) versus 6.2% with T2DM (STEP 2 study) at 68 weeks^{26,27}. Furthermore, the dosage of GLP-1 RA medications in this study were likely for diabetes control and were not optimized for weight reduction.

Second, there is a risk of selection bias. However, rigorous selection criteria were applied to include only relatively healthy nonsurgical patients who could have been eligible for metabolic surgery if this option had been offered at the index date. Third, an ideal target trial emulation design to compare the outcomes with metabolic surgery versus GLP-1 RAs would require that the study baseline characteristics be captured and the follow-up to begin at the time that the decision is made for an individual to undergo metabolic surgery versus receiving GLP-1 RAs because it is at this point that care for the individual and their interaction with the healthcare system changes, depending on the decision^{29–31}. Such information was not available in our dataset, limiting emulation for the ideal target trial. However, based on the findings from the main and sensitivity analyses, our inability to emulate an ideal target trial is unlikely to have had a meaningful impact on the clinical interpretation of the study results. Fourth, although overlap weighting created a precise balance for a long list of important covariates^{14,15}, and regression adjustment was used to control for confounding, unmeasured confounders (at baseline and during follow-up) may have influenced the findings of this observational study, meaning causal inference cannot be assumed. The E-value can help assess the potential impact of unmeasured confounding factors³². The magnitude of the associations between known risk factors and the study endpoints was smaller than the estimated E-values for mortality, MACE, nephropathy and retinopathy. This suggests that unknown or unmeasured confounders are unlikely to negate the observed favorable association between metabolic surgery and the study endpoints. In addition, the Kaplan–Meier curves show a higher initial incidence of the studied events after metabolic surgery, followed by a parallel trend over several years before the curves diverge (Fig. 1). If the separation of curves favoring metabolic surgery had occurred immediately after the index dates, it would have strongly suggested the presence of selection bias and residual confounding.

In conclusion, given the large sample size, using doubly robust statistical estimation, and a long follow-up time, the M6 study provides strong evidence that among patients with T2DM and obesity, metabolic surgery, compared with GLP-1 RA medications, is associated with a lower risk of incident macrovascular and microvascular morbidity and mortality. In the absence of RCTs, the current study represents the most reliable evidence that metabolic surgery is superior to medical treatment even with the availability of GLP-1RAs. Future studies should compare the cardiometabolic outcomes of metabolic surgery with newer GLP-1RAs that are more effective for weight reduction.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-025-03893-3>.

References

1. Courcoulas, A. P. et al. Long-term outcomes of medical management vs bariatric surgery in type 2 diabetes. *JAMA* **331**, 654–664 (2024).
2. Mingrone, G. et al. Bariatric–metabolic surgery versus conventional medical treatment in obese patients with type 2 diabetes: 5 year follow-up of an open-label, single-centre, randomised controlled trial. *Lancet* **386**, 964–973 (2015).
3. Ikramuddin, S. et al. Lifestyle intervention and medical management with vs without Roux-en-Y gastric bypass and control of hemoglobin A_{1c}, LDL cholesterol, and systolic blood pressure at 5 years in the diabetes surgery study. *JAMA* **319**, 266–278 (2018).
4. Aminian, A. et al. Association of metabolic surgery with major adverse cardiovascular outcomes in patients with type 2 diabetes and obesity. *JAMA* **322**, 1271–1282 (2019).

5. Fisher, D. P. et al. Association between bariatric surgery and macrovascular disease outcomes in patients with type 2 diabetes and severe obesity. *JAMA* **320**, 1570–1582 (2018).
6. O'Brien, R. et al. Microvascular outcomes in patients with diabetes after bariatric surgery versus usual care: a matched cohort study. *Ann. Intern. Med.* **169**, 300–310 (2018).
7. Wolff Sagy, Y. et al. Effectiveness of bariatric metabolic surgery versus glucagon-like peptide-1 receptor agonists for prevention of congestive heart failure. *Nat. Med.* **30**, 2337–2342 (2024).
8. American Diabetes Association. 10. Cardiovascular disease and risk management: *Standards of Medical Care in Diabetes—2020*. *Diabetes Care* **43**, S111–S134 (2020).
9. Marso, S. P. et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N. Engl. J. Med.* **375**, 311–322 (2016).
10. Marso, S. P. et al. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N. Engl. J. Med.* **375**, 1834–1844 (2016).
11. Gerstein, H. C. et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet* **394**, 121–130 (2019).
12. Lincoff, A. M. et al. Semaglutide and cardiovascular outcomes in obesity without diabetes. *N. Engl. J. Med.* **389**, 2221–2232 (2023).
13. Stenberg, E. & Näslund, E. Major adverse cardiovascular events among patients with type-2 diabetes, a nationwide cohort study comparing primary metabolic and bariatric surgery to GLP-1 receptor agonist treatment. *Int. J. Obes.* **47**, 251–256 (2023).
14. Thomas, L. E., Li, F. & Pencina, M. J. Overlap weighting: a propensity score method that mimics attributes of a randomized clinical trial. *JAMA* **323**, 2417–2418 (2020).
15. Funk, M. J. et al. Doubly robust estimation of causal effects. *Am. J. Epidemiol.* **173**, 761–767 (2011).
16. Pi-Sunyer, X. et al. A randomized, controlled trial of 3.0 mg of liraglutide in weight management. *N. Engl. J. Med.* **373**, 11–22 (2015).
17. Trujillo, J. M., Nuffer, W. & Smith, B. A. GLP-1 receptor agonists: an updated review of head-to-head clinical studies. *Ther. Adv. Endocrinol. Metab.* **12**, 204201882199732 (2021).
18. Douros, J. D., Tong, J. & D'Alessio, D. A. The effects of bariatric surgery on islet function, insulin secretion, and glucose control. *Endocr. Rev.* **40**, 1394–1423 (2019).
19. Batterham, R. L. & Cummings, D. E. Mechanisms of diabetes improvement following bariatric/metabolic surgery. *Diabetes Care* **39**, 893–901 (2016).
20. Gasoyan, H., Pföh, E. R., Schulte, R., Le, P. & Rothberg, M. B. Early- and later-stage persistence with antiobesity medications: a retrospective cohort study. *Obesity* **32**, 486–493 (2024).
21. Liss, D. T. et al. Treatment modification after initiating second-line medication for type 2 diabetes. *Am. J. Manag. Care* **29**, 661–668 (2023).
22. Wilding, J. P. H. et al. Weight regain and cardiometabolic effects after withdrawal of semaglutide: the STEP 1 trial extension. *Diabetes Obes. Metab.* **24**, 1553–1564 (2022).
23. Bashir, B. et al. Microvascular complications of obesity and diabetes—role of bariatric surgery. *Obes. Rev.* **24**, e13602 (2023).
24. Vivante, A. et al. Body mass index in 1.2 million adolescents and risk for end-stage renal disease. *Arch. Intern. Med.* **172**, 1644–1650 (2012).
25. Jastreboff, A. M. et al. Tirzepatide once weekly for the treatment of obesity. *N. Engl. J. Med.* **387**, 205–216 (2022).
26. Davies, M. et al. Semaglutide 2·4 mg once a week in adults with overweight or obesity, and type 2 diabetes (STEP 2): a randomised, double-blind, double-dummy, placebo-controlled, phase 3 trial. *Lancet* **397**, 971–984 (2021).
27. Wilding, J. P. H. et al. Once-weekly semaglutide in adults with overweight or obesity. *N. Engl. J. Med.* **384**, 989–1002 (2021).
28. Gasoyan, H. et al. One-year weight reduction with semaglutide or liraglutide in clinical practice. *JAMA Netw. Open* **7**, e243326 (2024).
29. Hernán, M. A., Wang, W. & Leaf, D. E. Target trial emulation: a framework for causal inference from observational data. *JAMA* **328**, 2446–2447 (2022).
30. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
31. Madenci, A. L. et al. Estimating the effect of bariatric surgery on cardiovascular events using observational data? *Epidemiology* **35**, 721–729 (2024).
32. VanderWeele, T. J. & Ding, P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
33. Inker, L. A. et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N. Engl. J. Med.* **385**, 1737–1749 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Hamlet Gasoyan  ¹, **Mohammad Hesam Alavi** ², **Alexander Zajichek** ³, **Nicholas J. Casacchia**  ¹, **Abdullah Al Jabri**  ², **James Bena**  ³, **Xiaoxi Feng** ², **Rickesha Wilson** ², **Ricard Corcelles** ², **W. Scott Butsch** ², **Rishi P. Singh** ⁴, **Nikhil Das** ⁵, **Hejin Jeong**  ⁶, **Amgad Mentias** ⁷, **W. H. Wilson Tang**  ⁷, **Bartolome Burguera** ⁸, **Raul J. Rosenthal** ⁹, **Steven E. Nissen**  ⁷✉, **Michael B. Rothberg**  ¹ & **Ali Aminian**  ²✉

¹Center for Value-Based Care Research, Department of Internal Medicine and Geriatrics, Primary Care Institute, Cleveland Clinic, Cleveland, OH, USA.

²Bariatric and Metabolic Institute, Department of General Surgery, Cleveland Clinic, Cleveland, OH, USA. ³Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA.

⁴Department of Ophthalmology, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ⁵Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ⁶Case Western Reserve University School of Medicine, Cleveland, OH, USA. ⁷Heart, Vascular and Thoracic Institute, Cleveland Clinic, Cleveland, OH, USA. ⁸Endocrinology and Metabolism Institute, Cleveland Clinic, Cleveland, OH, USA. ⁹Bariatric and Metabolic Institute, Department of General Surgery, Cleveland Clinic, Weston, FL, USA.

✉ e-mail: nissens@ccf.org; aminiaa@ccf.org

Methods

In this retrospective study, the target population included patients with T2DM and obesity who were managed at the CCHS between 2010 and 2017. Patients were eligible to receive any diabetes medications, including GLP-1 RAs. In this observational analog of the intention-to-treat analysis, patients who underwent metabolic surgery during this period were compared with those who met the eligibility criteria for metabolic surgery on their assigned index dates but did not undergo surgery. In the nonsurgical group, we only included patients receiving medical management with GLP-1 RA medications. This was defined as having three or more documented prescription fills in 1 year before their assigned index date and three or more fills in 1 year after their assigned index date (see the 'Study cohorts and enrollment criteria' section).

The study used the Cleveland Clinic electronic health records (EHRs), including Surescripts dispensed prescription records³⁴, through 31 December 2022. The International Classification of Diseases and Current Procedural Terminology procedure codes used in this study³⁵ are provided in Supplementary Table 3. The study outcomes during the follow-up were ascertained by reviewing notes and tests in the EHRs, either from internal CCHS encounter data or through the Epic Care Everywhere interoperability platform, which provides exchange of EHR data between participating healthcare systems.

The study was registered on the ClinicalTrials.gov website (registration no. [NCT06355219](#)). The study was approved by the CCHS institutional review board as minimal-risk research using data collected for routine clinical practice, for which the requirement for informed consent was waived (study ID 19-066). The Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines were followed³⁶.

Study cohorts and enrollment criteria

Adult patients (aged between 18 and 75 years) with obesity (defined by a BMI $\geq 30 \text{ kg m}^{-2}$) and T2DM (defined by either HbA1c $\geq 6.5\%$ or by taking at least one diabetes medication) who underwent RYGB or sleeve gastrectomy at CCHS locations in Florida and Ohio between 1 January 2010 and 31 December 2017, were identified for inclusion in the metabolic surgery group. Patients with a history of organ transplantation, cardiac ejection fraction lower than 20% and malignancies were excluded (Extended Data Fig. 1). The date of first metabolic surgery was used as an index date for surgical patients.

A stringent process was followed to identify patients in the non-surgical group. In the large pool of CCHS patients with a diagnosis of T2DM (between 1 January 2010 and 31 December 2017) who did not have a history of metabolic surgery, five dates from the collection of index dates of surgical patients were randomly assigned to each patient, which served as potential index dates for nonsurgical patients. On each assigned index date, nonsurgical patients who met the age, BMI, T2DM and the other aforementioned selection criteria for surgical patients were considered for enrollment. Patients were also required to receive GLP-1 RA medications on the assigned index date, defined by a prescription order for GLP-1RA placed between 2010 and 2017, as well as three or more documented prescription fills within 1 year before their assigned index date and three or more fills within 1 year after their assigned index date. We then selected the earliest eligible index date (from the available five potential index dates per patient) for each included nonsurgical patient. This approach allowed us to construct a reasonable comparator group to the metabolic surgery group from a sample of nonsurgical patients receiving GLP-1RA medications who could have been potentially eligible for metabolic surgery if that option had been offered to them at the time of the assigned index date.

Primary and secondary endpoints

The primary endpoint was all-cause mortality. Death information was obtained from a combination of the EHR, social security data and

state death indices. Secondary endpoints included incident MACE, nephropathy and retinopathy.

MACE was a composite of four adverse outcomes, defined as the first occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation (Extended Data Table 1). The earliest occurrence of a coronary artery event, cerebrovascular event, heart failure or atrial fibrillation after the index date was recorded as the MACE event date. If a patient had history of one of these conditions before the index date, a new event during the follow-up was considered based on having either surgery or a procedure for that condition, hospitalizations or laboratory values indicative of exacerbation of the condition. The details are explained in Extended Data Table 1.

Incident nephropathy was defined as onset of 40% or more sustained decline in eGFR compared with baseline (calculated using the 2021CKD-EPI creatinine equation)³³, onset of sustained eGFR less than 15 ml min 1.73 m^{-2} , initiation of dialysis or kidney transplant.

Incident retinopathy was defined by (1) the first occurrence of any retinopathy in patients who did not have retinopathy at baseline or (2) progression to a more severe form of retinopathy in those with baseline retinopathy. The later included (2a) progression from mild or moderate nonproliferative retinopathy to severe nonproliferative retinopathy after the index date or (2b) progression from nonproliferative retinopathy to proliferative retinopathy after the index date (Extended Data Table 3).

Other outcomes

Weight, HbA1c levels and the dates of prescription orders and dispenses of medications for T2DM and cardiovascular conditions were collected using EHR data linked to Surescripts dispensing records³⁴, to compare patients in the metabolic surgery and GLP-1 RA groups over time; 90-day complications of metabolic surgery were collected.

Statistical analysis

The study used doubly robust estimation combining overlap weighting and outcome regression to minimize the effects of confounding factors when comparing outcomes in the metabolic surgery group and the GLP-1RA group^{14,15}. Thirty-three a priori-identified potential confounders (including sex, age at the index date, race, CCHS hospital location (Florida versus Ohio), household income estimate based on zip code, smoking status and BMI; clinical and laboratory data on systolic BP, triglycerides, HDL-C, LDL-C, HbA1c, eGFR and UACR; medical history of hypertension, dyslipidemia, peripheral neuropathy, heart failure, myocardial infarction, coronary artery disease, chronic obstructive pulmonary disease, atrial fibrillation, peripheral arterial disease, cerebrovascular events, nephropathy and retinopathy; and medication history of non-insulin diabetes drugs, insulin, lipid-lowering medications, RAAS inhibitors and other antihypertensive medications, aspirin and warfarin) were used for overlap weighting. Next, in the analytical phase, the same variables used for overlap weighting, with the addition of the index date, were controlled in the weighted fully adjusted Cox proportional hazards regression model with robust standard errors (sandwich estimators) for the regression coefficients.

To address the numerical variables with missing values at baseline, multiple imputation using chained equations was used to create five imputed datasets using predictive mean matching. Rubin's formula was used to derive imputation-corrected standard errors of model estimates and contrasts^{37,38}. Missing values for race and smoking status variables were coded as a separate category and included in the analyses as such.

For all-cause mortality, MACE, nephropathy and retinopathy, the cumulative incidence estimates (using the Kaplan–Meier method), unadjusted absolute risk differences (weighted by the overlap weights) and adjusted absolute risk differences (using fully adjusted Cox models)³⁹ were calculated for 10 years after the index date. The 95% CIs for the difference in the 10-year risk were obtained by combining

percentiles across the bootstrap distributions from all five imputation datasets.

A flexible random-intercept linear model, with four-knot restricted cubic spline for time \times treatment interaction on measurements was used to compare mean changes in weight and HbA1c during the study follow-up. Point estimates were averaged over the imputation datasets. The 95% CIs were estimated from patient-level bootstrap resampling. The 2.5th and 97.5th percentiles were taken from the combined set of bootstrap estimates across all imputations.

A two-sample proportions test was used to compare the proportions of patients with active prescriptions and fills of medications for T2DM and cardiovascular disease between the study groups at 10 years of follow-up after the index dates. Point estimates were averaged over the imputation datasets. Pooled standard errors were obtained using Rubin's formula³⁸.

Given the nature of the study, all findings should be considered exploratory. All analyses were performed using R v.4.2.1.

Sensitivity analyses

The primary analysis plan for this study was intention-to-treat, reflecting the real-world scenario where not all participants may adhere perfectly to the treatment during follow-up. In the primary analysis, nonsurgical patients who were nonadherent to GLP-1RA medications after baseline were not censored. However, to address noncompliance in the nonsurgical group, we performed two sensitivity analyses. In the first sensitivity analysis, nonsurgical patients who discontinued GLP-1RA medications (defined by a gap of more than 6 months in their active prescription orders/dispensing records)⁴⁰ were censored at the earliest occurrence of such events. Furthermore, we conducted an additional sensitivity analysis modeling an extreme hypothetical scenario in which nonsurgical patients who discontinued GLP-1RA medications remained event-free throughout their entire follow-up (representing the best-case scenario for nonsurgical patients). In that analysis, the main study outcomes (mortality, MACE, nephropathy and retinopathy) were not counted for nonsurgical patients who discontinued GLP-1RA medications from the discontinuation date until the end of the available follow-up.

We conducted a third sensitivity analysis where surgical patients who continued to receive GLP-1RAs after surgery or initiated GLP-1RAs during follow-up were censored at the earliest occurrence of the medication prescription order.

Furthermore, to assess the robustness of the identified associations between metabolic surgery and the study endpoints to potentially unmeasured confounders, E-values were calculated³². As explained by VanderWeele & Ding³², the E-value represents the minimum strength of association—on the risk ratio scale—that an unmeasured confounder would need to have with both the treatment and the outcome, conditional on the measured covariates, to fully explain away a specific treatment–outcome association. A small E-value indicates that only a weak unmeasured confounder would be needed to negate the observed effect estimate. E-values (for both the HR estimates and the upper limit of their 95% CIs) were calculated for the study outcomes, along with the HRs of known confounders (for example, hypertension, insulin use).

Given prior concerns that observational efficacy estimates of metabolic surgery may be overstated because of positive changes during the preoperative period associated with medical screening and optimization³¹, we compared patients' health status between the study index date and either the referral date for metabolic surgery (in the surgical group) or the GLP-1RA initiation date (in the nonsurgical group).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset generated during the current study is not publicly available to protect patient confidentiality. However, it is available to academic

investigators upon request, pending the receipt of a signed data sharing agreement and review of the study protocol (approved by a local institutional review board or research ethics committee), statistical analysis plan and publication plan. All data sharing requests must be approved by the CCHS institutional review board and the Law Department before de-identified data can be shared. The corresponding authors will respond to requests within 2 months of receipt.

References

34. Blecker, S. et al. Validation of EHR medication fill data obtained through electronic linkage with pharmacies. *J. Manag. Care Spec. Pharm.* **27**, 1482–1487 (2021).
35. Milinovich, A. & Kattan, M. W. Extracting and utilizing electronic health data from Epic for research. *Ann. Transl. Med.* **6**, 42 (2018).
36. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J. Clin. Epidemiol.* **61**, 344–349 (2008).
37. Van Buuren, S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **16**, 219–242 (2007).
38. Rubin, D. B. Inference and missing data. *Biometrika* **63**, 581–592 (1976).
39. Austin, P. C. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J. Clin. Epidemiol.* **63**, 46–55 (2010).
40. Raebel, M. A., Schmitttdiel, J., Karter, A. J., Konieczny, J. L. & Steiner, J. F. Standardizing terminology and definitions of medication adherence and persistence in research employing electronic databases. *Med. Care* **51**, S11–S21 (2013).

Acknowledgements

We thank M. E. Satava from the Cleveland Clinic, for her help in collecting some of the data. She did not receive additional compensation, outside her usual salary, for her contribution.

Author contributions

H.G., S.E.N., M.B.R. and A.A. contributed to study conception and design. H.G., M.H.A., N.J.C., A.A.J., N.D., H.J. and A.A. were involved in data acquisition. A.Z. performed the statistical analysis. All authors contributed to the interpretation of the data. H.G. and A.A. prepared the first draft of the manuscript. All authors critically revised the manuscript for intellectual content and clarity, and approved the final version for submission. A.A. provided administrative support and supervised the work.

Competing interests

W.S.B. has received honoraria and paid consultancy from Novo Nordisk, Abbott Nutrition, Medscape, Alfie Health and the Med Learning Group. R.P.S. reported receiving personal fees from Alcon, Apellis, Bausch + Lomb, EyePoint, Genentech, Iveric Bio, Regeneron, Regenxbio and ZEISS, and research grants from Janssen. W.H.W.T. received consulting fees from Sequana Medical, Cardiol Therapeutics, Genomics plc, Zehna Therapeutics, WhiteSwell, Boston Scientific, CardiaTec Biosciences, Intellia Therapeutics, Bristol Myers Squibb, Alleviant Medical, Alexion Pharmaceuticals, Salubris Biotherapeutics and BioCardia, and received honoraria from Springer, Belvoir Media Group and the American Board of Internal Medicine. B.B. received an honorarium from Novo Nordisk. R.J.R. reported receiving personal fee from Medtronic, Diagnostic Green, mediCAD and Dendrite Imaging, and serving as the CEO of Dendrite Imaging. S.E.N. received grants to perform clinical trials from AbbVie, AstraZeneca, Amgen, Bristol Myers Squibb, Eli Lilly, Esperion Therapeutics, Medtronic, MyoKardia, New Amsterdam Pharmaceuticals, Novartis and Silence Therapeutics. M.B.R. has a consulting relationship with the Blue Cross Blue Shield

Association. A.A. received research grants from Medtronic and Ethicon, and serves as a consultant for Medtronic, Ethicon and Eli Lilly. The other authors declare no competing interests.

Additional information

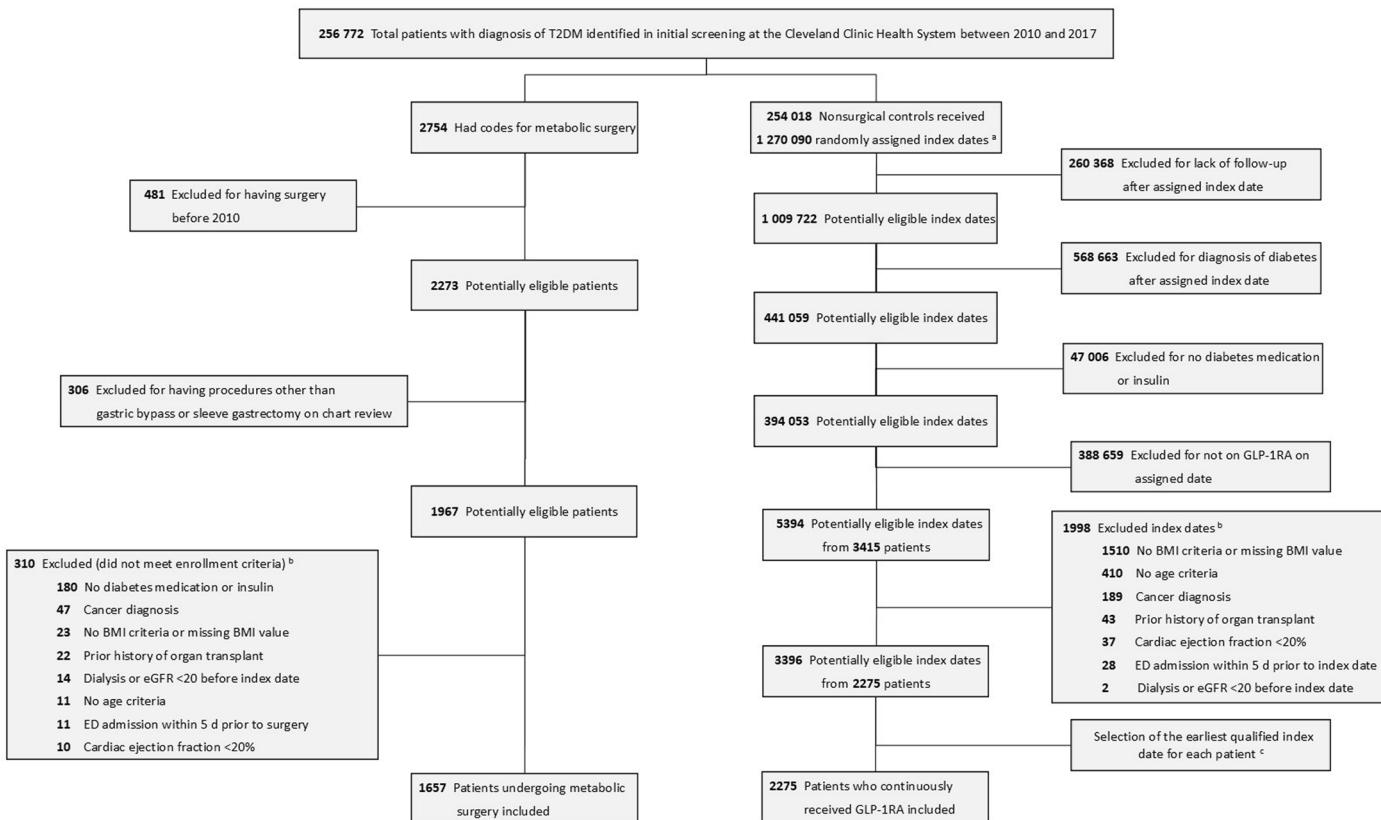
Extended data is available for this paper at
<https://doi.org/10.1038/s41591-025-03893-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-025-03893-3>.

Correspondence and requests for materials should be addressed to Steven E. Nissen or Ali Aminian.

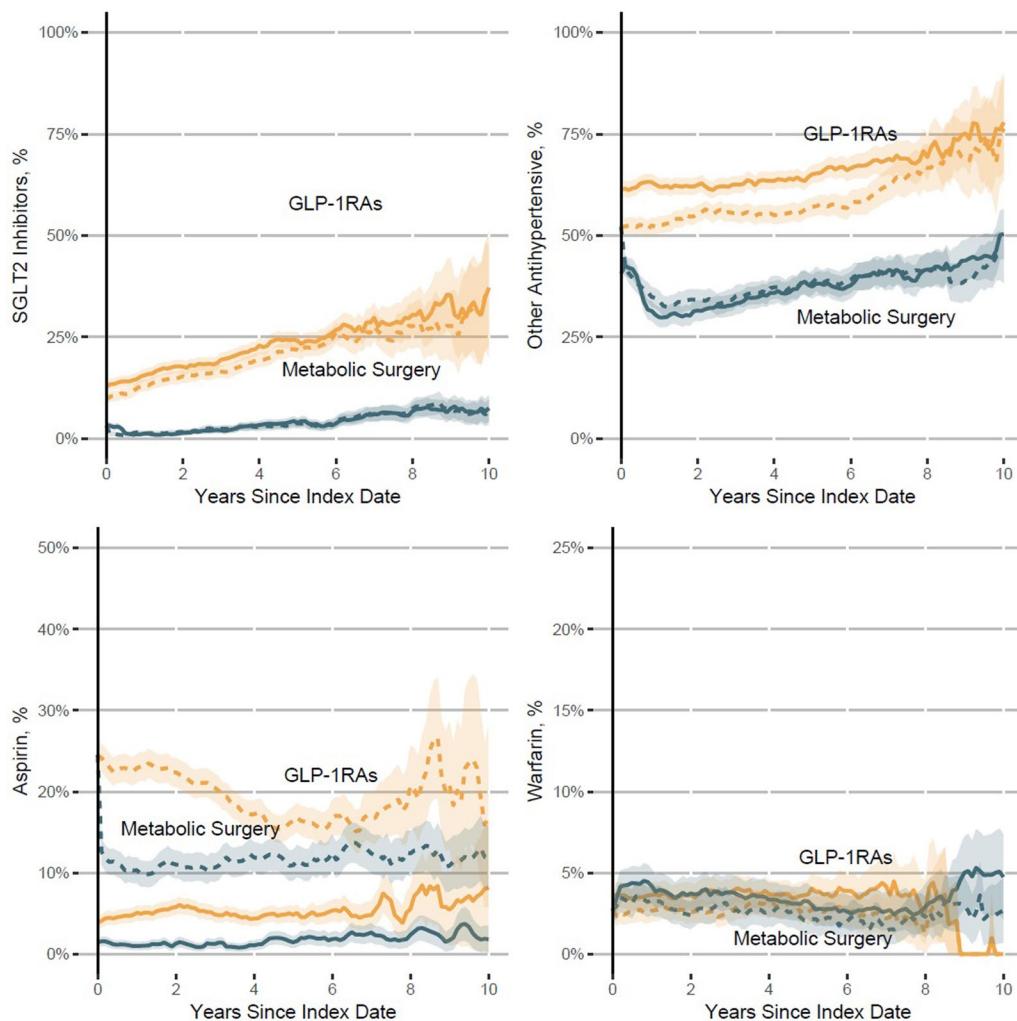
Peer review information *Nature Medicine* thanks Anne Ehlers, Abdelrahmin Nimeri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Michael Basson, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Identification of eligible patients for inclusion. Details of International Classification of Diseases (ICD) and Current Procedural Terminology (CPT) codes used in cohort construction are available in Supplementary Table 3.^a Five dates from the collection of index dates of surgical patients were randomly assigned to each nonsurgical control, which served as potential index dates for nonsurgical patients.^b Some patients met multiple

exclusion criteria.^c In case of more than one index date that a patient would be eligible for inclusion, we selected the earliest eligible index date for each nonsurgical patient. Abbreviations: BMI, body mass index; ED, emergency department; eGFR, estimated glomerular filtration rate; GLP-1RA · glucagon-like peptide-1 receptor agonist; T2DM, type 2 diabetes mellitus.



Extended Data Fig. 2 | Proportions of patients with orders and dispenses for other medications. Plots display proportions of patients (with 95% point-wise confidence intervals (shaded areas)) in the metabolic surgery and GLP-1RA groups over time who were prescribed or dispensed with the indicated medications. For a given time point (e.g., 1 year after the index date), prescription orders (dashed lines) were defined as the percentage of patients (of those who were followed up to at least that date) who had at least one prescription order

for that medication, with start and end dates that encompassed that time point (using the last follow-up date if the end date was missing). For a given time point (e.g., 1 year after the index date), medication dispenses (solid lines) were defined as the percentage of patients (of those who were followed up to at least that date) who had at least one dispense record for that medication in the preceding 6 months. Medication dispensation data was based on Surescripts data. SGLT2, Sodium-Glucose Cotransporter 2.

Extended Data Table 1 | Definition of major adverse cardiovascular events (MACE)

	In patients with no history at baseline	In patients with history of MACE before the index date
Coronary artery events	First occurrence of diagnostic/procedure code for unstable angina, myocardial infarction, or coronary intervention/surgery in patients who did not have history of coronary artery events at baseline, verified by a record of two additional codes to assure the accuracy of the event during follow-up.	a) procedure code for coronary intervention or coronary artery bypass graft surgery (CABG) during follow-up, or b) manual chart review in patients with hospitalization and laboratory evidence of elevated troponin levels (new acute myocardial infarction) during follow-up.
Cerebrovascular events	First occurrence of diagnostic/procedure code for ischemic stroke, hemorrhagic stroke, or carotid intervention/surgery in patients who did not have history of cerebrovascular events at baseline, verified by a record of two additional codes to assure the accuracy of the event during follow-up.	a) procedure code for carotid intervention or carotid surgery during follow-up or b) manual chart review in patients with hospitalization and principal/primary diagnosis of new stroke during follow-up.
Heart failure	First occurrence of diagnostic code for heart failure in patients who did not have history of heart failure at baseline, verified by a record of two additional codes to assure the accuracy of the event during follow-up.	Manual chart review in patients with at least 2 emergency department visits or a hospitalization with a diagnosis code for heart failure and laboratory evidence of elevated BNP levels.
Atrial fibrillation	First occurrence of diagnostic code for atrial fibrillation in patients who did not have history of atrial fibrillation at baseline, verified by a record of two additional codes to assure the accuracy of the event during follow-up.	Manual chart review in patients with an emergency department visit or hospitalization with principal/primary diagnosis of atrial fibrillation.

Diagnostic codes used in this study are presented in Supplementary Table 3. MACE was a composite end point, defined as the occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation during follow-up. The earliest occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation after the index date was recorded as the MACE event date.

Extended Data Table 2 | Incidence estimates, absolute risk differences and hazard ratios for individual components of MACE for metabolic surgery patients versus nonsurgical patients

Outcome	Metabolic Surgery (n=1657)			GLP-1RA (n=2275)			Unadjusted absolute 10-year risk difference (%) with 95% CI ^b	Adjusted absolute 10-year risk difference (%) with 95% CI ^c	Adjusted hazard ratio (95% CI) ^d	P value ^d
	N with Event	Rate per 100 patient- years ^a	Cumulative incidence (%) with 95% CI at 10 years ^b	N with Event	Rate per 100 patient- years ^a	Cumulative incidence (%) with 95% CI at 10 years ^b				
Heart Failure	81	0.84	9.4% (7.0%, 11.6%)	202	1.53	12.7% (9.9%, 15.6%)	3.3% (-0.3%, 7.5%)	8.7% (4.5%, 12.5%)	0.44 (0.3, 0.65)	<0.001
Coronary Artery Events	96	1.16	11.5% (8.6%, 14.6%)	233	1.48	11.5% (8.9%, 13.8%)	0.01% (-3.0%, 3.4%)	4% (0.2%, 7.5%)	0.7 (0.49, 0.98)	0.037
Cerebrovascular Events	31	0.36	4.0% (2.1%, 5.8%)	55	0.41	8.3% (2.9%, 19.8%)	4.3% (-1.5%, 15.7%)	0.1% (-3.7%, 3.6%)	0.98 (0.5, 1.9)	0.94
Atrial Fibrillation	106	1.24	11.9% (9.2%, 14.0%)	196	1.28	14.4% (10.2%, 18.7%)	2.5% (-3.0%, 7.6%)	1.7% (-2.3%, 5.6%)	0.84 (0.57, 1.24)	0.38

^aBased on overlap weighted data by taking the total number of (weighted) events observed divided by the total (weighted) event time (for those with events) or follow-up time (for those censored). ^bBased on overlap weighted data. ^cEstimated using fully adjusted Cox models. ^dHazard ratios (95% CIs) and P values from fully adjusted Cox models comparing the relative instantaneous risk of each outcome for surgical versus nonsurgical patients; the confidence intervals and two-sided P values were computed with a normal approximation. The exact P value for heart failure was 5.3×10^{-5} .

Extended Data Table 3 | Definitions of nephropathy and retinopathy

Outcome	Definition
Incident nephropathy	Time to the first occurrence of any component of the composite endpoint: <ul style="list-style-type: none"> • $\geq 40\%$ sustained decline in eGFR • End-stage renal disease, defined as the onset of sustained eGFR $< 15 \text{ mL/min/1.73 m}^2$, or the initiation of dialysis, or kidney transplant.
Development of new retinopathy during follow-up	First occurrence of diagnostic code for any retinopathy condition or related procedure code in patients who did not have history of retinopathy before the index date (as described in Supplementary Table 3), followed by a manual chart review to assure the accuracy of the events.
Progression of retinopathy during follow-up, diagnosed prior to index date	<ul style="list-style-type: none"> • Progression from mild/moderate before index date to severe NPDR after index date.^a • Progression from NPDR before index date to PDR after index date.^a

Comprehensive chart review of patients assessing ophthalmology encounters, notes, procedures, and diagnostic codes in EHR and scanned documents was performed. ^aWorsening of retinopathy was based on: (1) documentation by the ophthalmologist based on dilated retinal examination; (2) new macular edema, vitreous hemorrhage; (3) need for procedures including laser/panretinal photocoagulation (PRP), pars plana vitrectomy (PPV); (4) need for intravitreal injection (Avastin/bevacizumab) with no history of injection before the index date. Abbreviations: eGFR, estimated glomerular filtration rate; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.

Extended Data Table 4 | Incidence estimates and hazard ratios for macrovascular and microvascular outcomes for metabolic surgery patients versus nonsurgical patients, stratified according to surgical procedure

Outcome	Cumulative incidence (%) with 95% CI at 10 years ^a			Adjusted hazard ratio (95% CI) ^b	P Value ^c
	RYGB (n=1015)	SG (n=642)	GLP-1RA (n=2275)		
All-cause mortality	8.6 (5.9, 10.9)	9.1 (6.0, 12.8)	12.4 (9.9, 15.2)		
RYGB vs SG				1.15 (0.64, 2.07)	0.631
RYGB vs GLP-1RA				0.72 (0.49, 1.05)	0.086
SG vs GLP-1RA				0.62 (0.36, 1.07)	0.086
MACE ^d	21.0 (17.1, 25.1)	30.5 (22.5, 37.0)	34.0 (28.1, 44.2)		
RYGB vs SG				0.77 (0.54, 1.08)	0.127
RYGB vs GLP-1RA				0.58 (0.44, 0.77)	<0.001
SG vs GLP-1RA				0.76 (0.56, 1.03)	0.075
Nephropathy	18.2 (13.6, 21.8)	28.4 (21.8, 35.8)	37.0 (30.2, 45.1)		
RYGB vs SG				0.54 (0.38, 0.76)	<0.001
RYGB vs GLP-1RA				0.41 (0.31, 0.54)	<0.001
SG vs GLP-1RA				0.76 (0.57, 1.01)	0.061
Retinopathy	6.1 (4.3, 8.7)	3.7 (1.5, 6.3)	15.9 (11.2, 25.0)		
RYGB vs SG				1.25 (0.58, 2.73)	0.567
RYGB vs GLP-1RA				0.49 (0.29, 0.84)	0.009
SG vs GLP-1RA				0.39 (0.19, 0.81)	0.011

^aBased on overlap weighted data. ^bEstimated using fully adjusted Cox models. ^cHazard ratios (95% CIs) and P values from fully adjusted Cox models comparing the relative instantaneous risk of each outcome for surgical and nonsurgical patients; the confidence intervals and two-sided P values were computed with a normal approximation. ^dA composite end point that was defined as first occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation. The exact P values were: MACE (RYGB versus GLP-1RA)= 1.3×10^{-4} ; nephropathy (RYGB versus SG)= 4.8×10^{-4} ; and retinopathy (RYGB versus GLP-1RA)= 1.4×10^{-10} . Abbreviations: RYGB, Roux-en-Y gastric bypass; SG, sleeve gastrectomy.

Extended Data Table 5 | Differences in proportion of patients on diabetes and cardiovascular medications at 10 years from baseline between groups

Medication	Unweighted Data		Overlap Weighted Data	
	Estimate (95% CI) ^a	P-value ^b	Estimate (95% CI) ^a	P-value ^b
Oral hypoglycemic agents	50.2% (38%, 62.5%)	1.1 X 10 ⁻¹⁵	47.9% (35.8%, 60%)	8.4 X 10 ⁻¹⁵
Insulin	28.6% (14.3%, 42.9%)	8.8 X 10 ⁻⁵	29.6% (15.2%, 44%)	5.7 X 10 ⁻⁵
GLP-1RAs	61.8% (47.9%, 75.7%)	1.0 X 10 ⁻⁵⁰	61.6% (47.8%, 75.5%)	1.0 X 10 ⁻⁵⁰
Semaglutide or Tirzepatide	19.6% (7%, 32.1%)	0.002	17.8% (5.6%, 30%)	0.004
SGLT2 inhibitors	29% (14.7%, 43.2%)	6.9 X 10 ⁻⁵	27% (12.7%, 41.3%)	2.2 X 10 ⁻⁴
Lipid-lowering drugs	52.7% (42.3%, 63.1%)	1.0 X 10 ⁻⁵⁰	49.3% (39.5%, 59%)	1.0 X 10 ⁻⁵⁰
RAAS inhibitors ^c	39.7% (25.3%, 54%)	6.4 X 10 ⁻⁸	38.7% (24.1%, 53.3%)	2.1 X 10 ⁻⁷
Other antihypertensive drugs	32.1% (17.9%, 46.3%)	9.1 X 10 ⁻⁶	32% (18.1%, 46%)	7.0 X 10 ⁻⁶
Aspirin ^d	8.7% (-3.2%, 20.7%)	0.152	5.1% (-6.7%, 16.9%)	0.397
Warfarin ^d	-2.8% (-4.8%, -0.7%)	0.007	-2.7% (-4.7%, -0.7%)	0.008

^aEstimated between-group difference (GLP-1RA group—metabolic surgery) in the proportion of patients with prescription order for diabetes and cardiovascular medications at 10 years from baseline. Point estimates were averaged over imputation datasets. Standard errors were estimated from the normal-based pooling of independent proportions and pooled across imputations using Rubin's formula. ^bTwo-sided P values from normal quantiles from the pooled estimates are displayed. ^cRenin–angiotensin–aldosterone system (RAAS) inhibitors included angiotensin-converting enzyme inhibitors and angiotensin receptor blockers. ^dThe lack of positive effects of metabolic surgery on aspirin and warfarin use was reassuring. Abbreviations: GLP-1RAs, glucagon-like peptide-1 receptor agonists; SGLT2, sodium-glucose cotransporter 2.

Extended Data Table 6 | Results of the first and second sensitivity analyses from fully adjusted Cox models for each outcome for metabolic surgery patients versus nonsurgical patients

Outcome	Sensitivity Analysis 1		Sensitivity Analysis 2	
	<i>After censoring nonsurgical patients who discontinued GLP-1RA medications during follow-up at the earliest occurrence of such an event</i>		<i>After modeling for the extreme hypothetical scenario that nonsurgical patients who discontinued GLP-1RA medications remained event-free throughout their entire follow-up</i>	
	Adjusted hazard ratio (95% CI) ^a	P value ^a	Adjusted hazard ratio (95% CI) ^a	P value ^a
All-cause mortality	0.78 (0.52, 1.17)	0.23	1.06 (0.72, 1.55)	0.784
MACE ^b	0.64 (0.50, 0.82)	<0.001	0.79 (0.62, 1.00)	0.055
Nephropathy	0.52 (0.40, 0.67)	<0.001	0.69 (0.54, 0.87)	0.002
Retinopathy	0.48 (0.29, 0.78)	0.003	0.60 (0.37, 0.96)	0.034

In the sensitivity analysis 1, nonsurgical patients who discontinued GLP-1RA medications (defined by a gap of >6 months in their active prescription orders/dispensation records) were censored at the earliest occurrence of such events. In the sensitivity analysis 2, representing the best-case scenario for nonsurgical patients, nonsurgical patients who discontinued GLP-1RA medications were modeled to not have experienced the study's primary or secondary events from the discontinuation date until the end of the available follow-up. ^aHazard ratios (95% CIs) and P values from fully adjusted Cox models comparing the relative instantaneous risk of each outcome for surgical versus nonsurgical patients. ^bMajor adverse cardiovascular events (MACE) were a composite end point that was defined as the first occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation. In the sensitivity analysis 1, the exact P values for MACE=4.7×10⁻⁴ and for nephropathy=2.7×10⁻⁷.

Extended Data Table 7 | Results of the third sensitivity analysis from fully adjusted Cox models for each outcome for metabolic surgery patients versus nonsurgical patients

Outcome	Sensitivity Analysis 3	
	<i>After censoring surgical patients who continued to receive GLP-1RA after surgery or initiated the GLP-1RA during follow-up</i>	
	Adjusted hazard ratio (95% CI) ^a	P value ^a
All-cause mortality	0.72 (0.50, 1.03)	0.07
MACE ^b	0.68 (0.53, 0.87)	0.002
Nephropathy	0.54 (0.43, 0.69)	<0.001
Retinopathy	0.39 (0.23, 0.66)	<0.001

In the third sensitivity analysis, surgical patients who continued to receive GLP-1RA after surgery or initiated GLP-1RA during the follow-up were censored at the earliest occurrence of the medication prescription order. ^aHazard ratio (95% CIs) and P values from fully adjusted Cox models comparing the relative instantaneous risk of each outcome for surgical versus nonsurgical patients. ^bMajor adverse cardiovascular events (MACE) were a composite end point defined as the first occurrence of coronary artery events, cerebrovascular events, heart failure or atrial fibrillation. The exact P values for nephropathy=4.8×10⁻⁷ and for retinopathy=4.4×10⁻⁴.

Extended Data Table 8 | E-values for the association of metabolic surgery on study outcomes and their upper limits of 95% confidence intervals in fully adjusted Cox models

Outcome	E-value for HR estimate	E-value for upper limit of 95% CI	Variable	Level	HR (95% CI)
All-cause mortality	2.32	1.25	Smoking status	Former vs. Never	1.32 (0.94, 1.85)
				Current vs. Never	1.83 (0.89, 3.75)
			Hypertension	Yes vs. No	1.30 (0.74, 2.30)
			Dyslipidemia	Yes vs. No	0.88 (0.58, 1.33)
			Insulin Use	Yes vs. No	1.14 (0.79, 1.64)
			HbA1c (%)	≥7% vs. <7%	1.19 (0.83, 1.72)
Major adverse cardiovascular endpoints	2.03	1.73	Smoking status	Former vs. Never	1.02 (0.83, 1.25)
				Current vs. Never	0.88 (0.50, 1.54)
			Hypertension	Yes vs. No	0.97 (0.70, 1.35)
			Dyslipidemia	Yes vs. No	0.84 (0.66, 1.08)
			Insulin Use	Yes vs. No	0.91 (0.72, 1.15)
			HbA1c (%)	≥7% vs. <7%	1.09 (0.88, 1.36)
Nephropathy	2.50	2.40	Smoking status	Former vs. Never	0.89 (0.72, 1.10)
				Current vs. Never	1.45 (0.90, 2.36)
			Hypertension	Yes vs. No	1.33 (0.94, 1.89)
			Dyslipidemia	Yes vs. No	1.00 (0.77, 1.30)
			Insulin Use	Yes vs. No	1.12 (0.88, 1.41)
			HbA1c (%)	≥7% vs. <7%	1.03 (0.82, 1.29)
Retinopathy	3.74	2.01	Smoking status	Former vs. Never	0.74 (0.52, 1.05)
				Current vs. Never	0.80 (0.31, 2.10)
			Hypertension	Yes vs. No	1.98 (1.09, 3.60)
			Dyslipidemia	Yes vs. No	1.03 (0.64, 1.67)
			Insulin Use	Yes vs. No	1.66 (1.13, 2.43)
			HbA1c (%)	≥7% vs. <7%	2.34 (1.52, 3.62)

E-values were derived from pooled hazard ratio (HR) estimates. Hazard ratios were obtained from the doubly robust Cox proportional hazards models pooled over imputations. Approximations were used for MACE and nephropathy because the event rates were >15%, as recommended by VanderWeele & Ding.³² The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away a specific treatment-outcome association. A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. The E-values (both for the HR estimates and their upper limit of the 95% CI) of the study end points along with the HRs of known confounders (e.g. hypertension, insulin use) were calculated. These factors were added to the previously fit-adjusted model. In the sensitivity analysis, the same process was repeated, replacing these confounders specifically with risk factors (e.g. hypertension, insulin use); then, a full model with all variables was also fitted. Based on the calculated E-value for nephropathy, the observed HR of 0.52 could be explained away by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of 2.50-fold each, above and beyond the measured confounders, but weaker confounding could not do so; the confidence interval of HR could be moved to include the null by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of 2.40-fold each, above and beyond the measured confounders, but weaker confounding could not do so. The HRs of known risk factors for nephropathy were 1.33 for hypertension and 1.12 for insulin use. It is not likely that an unmeasured or unknown confounder would have a substantially larger effect on the incident nephropathy than these known risk factors by having a relative risk exceeding 2.50. In the sensitivity analysis, hazard ratios for all identified confounders remained below the E-value necessary to make the findings nonsignificant.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
 - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
 - Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Data were extracted from the electronic health records (EHRs) in the Cleveland Clinic Health System (CCHS) using the Structured Query Language (SQL) queries (SQL Server 2022).

Data analysis All analyses were performed using the R statistical programming software (version 4.2.1; Vienna, Austria).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The dataset generated during the current study is not publicly available in order to protect patient confidentiality. However, it is available to academic investigators upon request, pending the receipt of a signed data-sharing agreement and review of the study protocol (approved by a local institutional review board or research

ethics committee), statistical analysis plan, and publication plan. All data-sharing requests must be approved by the CCHS Institutional Review Board and the Law Department before de-identified data can be shared. The corresponding authors will respond to requests within two months of receipt.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex (biological attribute) was obtained from the electronic health records. Sex was included in the analyses because it could be associated with both exposure and study end points.

Reporting on race, ethnicity, or other socially relevant groupings

Race was obtained from the electronic health records and based on patients selecting from fixed categories.

Population characteristics

A total of 3932 adult patients (2307 [58.7%] female; 3076 [78.2%] White, 649 [16.5%] Black, 121 [3.1%] other races; mean age, 54.2 years [SD 11.0]; mean body mass index [BMI], 41.8 kg/m² [SD, 8.6]) with obesity and T2DM who received care at the Cleveland Clinic Health System in the United States from 2010 to 2017 were studied. There were 1657 patients in the metabolic surgery group and 2275 patients in the GLP-1RA group. The distribution of 33 a priori-identified baseline covariates for the metabolic surgery and the GLP-1RA group was precisely balanced (standardized differences = 0) after overlap weighting.

Recruitment

In this retrospective observational study, patients with obesity and type 2 diabetes who received their care at the Cleveland Clinic Health System and met the selection criteria were included.

There is a risk of selection bias. However, rigorous selection criteria were applied to include only relatively healthy nonsurgical patients who could have been eligible for metabolic surgery if this option had been offered at the index date.

Ethics oversight

This observational study was approved by the Cleveland Clinic institutional review board as minimal risk research.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

N=3932. No sample size calculation was performed in this observational retrospective study.

In retrospective chart reviews, the sample size is inherently determined by the number of eligible cases available within the study period. This study included all patients meeting inclusion criteria during the defined time frame, representing the complete cohort accessible for analysis.

Data exclusions

Adult patients (aged between 18 and 75 years) with obesity and T2DM who underwent Roux-en-Y gastric bypass or sleeve gastrectomy at Cleveland Clinic locations in Florida and Ohio between January 1, 2010, and December 31, 2017 were identified for inclusion in the metabolic surgery group. Patients with a history of organ transplantation, cardiac ejection fraction <20%, and malignancies were excluded. The date of first metabolic surgery was used as an index date for surgical patients.

A stringent process was followed to identify patients in the nonsurgical group. In the large pool of Cleveland Clinic patients with a diagnosis of T2DM (between January 1, 2010, and December 31, 2017) who did not have a history of metabolic surgery, five dates from the collection of index dates of surgical patients were randomly assigned to each patient, which served as potential index dates for nonsurgical patients. On each assigned index date, nonsurgical patients who met the age, BMI, T2DM, and other above-described selection criteria for surgical patients were considered for enrollment. Patients were also required to receive GLP-1RA medications, defined by prescription order for GLP-1RA placed between 2010 and 2017, as well as ≥ 3 documented prescription fills within 1 year before their assigned index date and ≥ 3 fills within 1 year after their assigned index date. We then selected the earliest eligible index date (from the available 5 potential index dates per patient) for each included nonsurgical patient. This approach allowed us to construct a reasonable comparator group to the metabolic surgery group from a sample of nonsurgical patients receiving GLP-1RA medications, who could have been potentially eligible for metabolic surgery if that option had been offered to them at the time of the assigned index date.

Replication

Although the primary analysis plan for this study was intention-to-treat, to address non-compliance in the nonsurgical group, we performed sensitivity analyses by censoring nonsurgical patients who discontinued GLP-1RA medications. The main findings of study were supported by these additional analyses.

Moreover, to assess the robustness of the identified associations, we censored surgical patients who continued to receive GLP-1RAs after metabolic surgery or initiated the GLP-1RAs during follow-up. The main findings of study were supported by these additional analyses.

Furthermore, to assess the robustness of the identified associations between metabolic surgery and the study end points to potentially unmeasured confounders, the E-values were calculated. The E-value represents the minimum strength of association—on the risk ratio scale—that an unmeasured confounder would need to have with both the treatment and the outcome, conditional on the measured covariates, to

fully explain away a specific treatment–outcome association. A small E-value indicates that only a weak unmeasured confounder would be needed to negate the observed effect estimate. The magnitude of the associations of the known risk factors with the study endpoints was smaller than the estimated E-values for all-cause mortality (2.32), MACE (2.03), nephropathy (2.50), and retinopathy (3.74), which makes it unlikely that unmeasured confounders could eliminate the favorable association between metabolic surgery and the study endpoints.

Randomization

This study is not a randomized trial. The study used doubly robust estimation combining overlap weighting and outcome regression to minimize the effects of confounding factors when comparing outcomes in the metabolic surgery group and the GLP-1RA group. Thirty-three priori identified potential confounders (including sex, age at index date, race, Cleveland Clinic hospital location [Florida versus Ohio], household income estimate based on zip code, smoking status, BMI; clinical and laboratory data on systolic blood pressure, triglycerides, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), HbA1c, eGFR, urine albumin-creatinine ratio (UACR); medical history of hypertension, dyslipidemia, peripheral neuropathy, heart failure, myocardial infarction, coronary artery disease, chronic obstructive pulmonary disease, atrial fibrillation, peripheral arterial disease, cerebrovascular events, nephropathy, retinopathy; as well as medication history of noninsulin diabetes drugs, insulin, lipid-lowering medications, renin-angiotensin-aldosterone system inhibitors, other antihypertensive medications, aspirin, and warfarin) were used for overlap weighting.

Blinding

Retrospective reviews rely on information that has already been documented in a non-blinded clinical setting.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

ClinicalTrials.gov ID NCT06355219

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

The study used Cleveland Clinic electronic health records (EHR), including Surescripts dispensed prescription records, through December 31, 2022. The diagnostic and procedure codes for study endpoints were available from both internal Cleveland Clinic encounter data and through the Epic Care Everywhere interoperability platform, which provides data exchange of EHRs between participating health care systems. Manual chart reviews were used for confirmation of diagnostic codes pertaining to primary and secondary study endpoints.

Outcomes

The primary endpoint was all-cause mortality. Secondary composite endpoints included incident major adverse cardiovascular events (MACE), nephropathy, and retinopathy. MACE was a composite of 4 adverse outcomes, defined as the first occurrence of coronary artery events, cerebrovascular events, heart failure, or atrial fibrillation. Incident nephropathy was defined as onset of $\geq 40\%$ sustained decline in eGFR compared with baseline (calculated using the 2021 CKD-EPI creatinine equation), onset of sustained eGFR $< 15 \text{ mL/min}/1.73 \text{ m}^2$, initiation of dialysis, or kidney transplant. Incident retinopathy was defined by (1) the first occurrence of any retinopathy in patients who did not have retinopathy at baseline or (2) progression to a more severe form of retinopathy in those with baseline retinopathy. The later included (2a) progression from mild/moderate nonproliferative retinopathy to severe nonproliferative retinopathy after the index date, or (2b) progression from nonproliferative retinopathy to proliferative retinopathy after the index date. Weight, HbA1c levels, and dates of prescription orders and dispenses of medications for T2DM and cardiovascular conditions were collected to compare the patients in the metabolic surgery and GLP-1RA groups over time. Ninety-day complications of metabolic surgery were collected.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

Authentication