Supplementary Online Content

Fisher DP, Johnson E, Haneuse S, et al. Association between bariatric surgery and macrovascular disease outcomes in patients with type 2 diabetes and severe obesity. *JAMA*. doi:10.1001/jama.2018.14619

eAppendix. Statistical Methods

eTable 1. Baseline Characteristics of Patients With Type 2 Diabetes Who Underwent Bariatric Surgery Who Were Included in the Study and Those That Were Excluded Due to Missing Data or Inability to Match to Nonsurgical Patients, 2005-2011

eTable 2. Reasons for Censoring for Surgical and Nonsurgical Populations

eTable 3. Cumulative Number of Incident Macrovascular Events and Deaths Among Bariatric Patients and Matched Nonsurgical Patients

eFigure. Sensitivity Analyses

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Statistical Methods

Overarching strategy

The study design was a matched cohort with surgical cases (i.e. the “exposed”) matched to non-cases (i.e. “unexposed”). In identifying matched, unexposed individuals, the most salient point is the need to define the start time, or “time zero” for the time-to-event analyses. For surgical cases, time zero was the date of surgery. Controls had to match by date requiring that they had to be enrolled in the health system on the date of the surgery for the surgical patient they were matched to. Once matched by date, two further matching strategies were considered. One was to construct a model predicting treatment allocation using all patients available for analysis during the study period using propensity scores to identify matched, unexposed individuals. The second strategy was to match directly on a list of a priori-specified covariates that might contribute to confounding bias. This was the approach used for the current study. Propensity score matching was not used because building a single regression model for treatment allocation that acknowledged heterogeneity across patients and across time would be challenging and may not yield an optimal match for important confounding factors. Moreover, the large amount of information available for analysis in this study facilitated matching directly on clinically important covariates, with a key benefit being that it ensures exact balance between the exposed and unexposed patients for many important confounders. A drawback of this matching strategy is that balance is not guaranteed for measured factors not included in the matching algorithm. Adjustment for these factors, however, was achieved by regression modeling.

Construction of analytic datasets

The following steps were used to construct the analytic data set:

- Identify all bariatric surgery cases that met the study inclusion/exclusion criteria as specified in the Methods section of the main article. This resulted in a sample of 6,291 surgical patients
- Of these, 952 were excluded because of missing pre-operative BMI, HbA1c, and/or serum creatinine measures in the 2 years before surgery, leaving 5,339 patients. See below for additional detail on missing data.
- Identify all non-surgical patients with at least one BMI ≥35 kg/m^2 who <u>did not</u> undergo bariatric surgery during the study period (N=320,345).
- For each bariatric surgery case:
    1. Identify a pool of potential controls who were enrolled at the time of the surgery and satisfied the study inclusion/exclusion criteria.
    2. Restrict the pool to patients who matched the bariatric case on the basis of: study site, gender, age (±10 years), BMI (±5 kg/m^2), HbA1c level (± absolute 1.5%) and insulin use.
    3. For each remaining control calculate the Mahalanobis distance with the bariatric patient on the basis of: age, BMI, HbA1c, diabetes duration, and the number of days of health care utilization in the 7-24 months prior to the date of surgery.

4. Controls were matched to only one case. In instances where a control was eligible for multiple surgery patients, they were assigned to the surgical patient with the lowest number of potential controls; potential control numbers were capped at 20 for these calculations. If two or more surgical patients had the same (lowest) number of controls available, the control in question was randomly assigned to one of the surgical patients.
5. Select up to 3 controls to be retained in the final matched cohort, using the controls with the smallest Mahalanobis distance.

- Note, a total of 38 surgical cases were unable to be matched; that is there were no non-surgical patients who matched these patients on the basis of site, age, BMI, insulin use, HbA1c, and gender.

Balance of confounder distributions

Covariate balance was assessed by calculating standardized differences between the baseline characteristics of the surgical and control patients (Table 1 of the article). Subsequent to this, further adjustment for all potential confounders, including those in the matching process, was achieved by using regression analysis. A detailed justification for including the covariates that were used in the matching process into the Cox regression models is provided by Sjolander and Greenland (Statistics in Medicine, 2013).

Statistical analysis

The primary outcome was a time-to-event phenomenon so Cox regression models were used to examine the association between bariatric surgery and incident macrovascular disease.

- Patients were followed from the index date (the date of bariatric surgery or, for non-surgical patients, the date of surgery for the patient for whom they were matched) until the first occurrence of either incident macrovascular disease or a censoring event (e.g., death, disenrollment from the health plans).
- Preliminary modeling showed that the proportional hazards assumption did not hold for bariatric surgery vs. usual care ($p < 0.001$ for an interaction with log-time in the Cox model). Consequently, modeling using a flexible time-varying hazard ratio association was pursued using restricted cubic splines with knots at the 5^{th}, 35^{th}, 65^{th} and 95^{th} percentiles of the observed follow-up time scale.
- Adjusted models were fit that included: age, race/ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, non-Hispanic other), surgical year, BMI, smoking status (current, former, never), duration of observed diabetes before surgery (defined as first observed diagnosis, laboratory value, or prescription indicating T2DM), insulin use, oral diabetes medication use, uncontrolled blood pressure (defined as either systolic ≥ 140 or diastolic ≥ 90 at two consecutive measures on different days), use of ACE-inhibitor, angiotensin receptor blocker (ARB) medications, use of any other antihypertensive medication, insurance type (commercial, Medicare, Medicaid, other), estimated glomerular filtration rate (eGFR), LDL cholesterol ≥2.59 mmol/L (≥100 mg/dL), serum triglycerides ≥1.7 mmol/L (≥150 mg/dL), use of cholesterol lowering medication (statin or other), and history of peripheral arterial disease, hypertension, dyslipidemia, or

microvascular disease (diabetic retinopathy, diabetic neuropathy, or diabetes with renal manifestations, end-stage renal disease, or dialysis) before surgery (defined based on ICD-9 and CPT-4 codes).

- Because of potential variation in care between health systems, study site (HP, KPSC, KPNC, and KPWA) was adjusted for by stratifying the baseline hazard function. The underlying model that was fit via partial likelihood was one for which the baseline hazard function was specific to the study site. Thus, no assumptions are made regarding between-site differences in risk (such as proportional hazards),, providing the most flexible approach to adjusting for site (at least within the context of a Cox model)

Cumulative incidence plots

Figure 2 in the article plots the cumulative incidence of macrovascular disease among the bariatric cases and controls for all macrovascular, cardiac, and cerebrovascular events and for those who died. These figures are not "unadjusted" as is typical for Kaplan-Meier estimates of incidence since the event rates presented were calculated from the matched cohort (i.e., the surgical patients and non-surgical patients used in the Kaplan-Meier calculations are already matched on site, gender, age, BMI, HbA1c level, insulin use, diabetes duration, and the number of days of health care utilization in the 7-24 months prior to the date of surgery). Because our sensitivity analysis comparing the matched adjusted Cox models and matched unadjusted Cox models (see Appendix Figure 4-A) yield nearly identical results, the cumulative incidence plots are a reasonable approximation of the underlying rates in the two groups.

Standard error estimation

Since all covariates used in the construction of the matched cohort were included in the model, no further statistical adjustment for the design was necessary to obtain valid standard error estimates.

Missing data

Missing data occurred at two points in the study. Of the 6,291 bariatric cases, 952 (15%) who met the inclusion/exclusion criteria had missing pre-operative BMI, HbA1c, and/or serum creatinine measures in the 2 years before surgery. The second source for missing data was within the final matched cohort where patients might have had some baseline covariates missing. Most commonly these were race/ethnicity and/or smoking status and/or elevated blood pressure and/or elevated triglyceride levels and/or elevated LDL levels.

Appendix Table 5 compares the baseline characteristics of the 15% of surgical patients who did not have either a pre-operative BMI or HbA1c and were lost during the matching process compared with patients who had complete pre-operative data. Surgical patients with missing data were similar for most characteristics. The characteristics that did differ (such as year of their surgical procedure and duration of observed diabetes) suggest that patients with missing data were either new to our health care systems (i.e., had not been receiving care long enough to have all of these baseline data captured) or had undergone their bariatric procedure outside of our integrated network. For example, these patients were more often from one of our health systems (HealthPartners: 35% of missing vs. 5% non-missing); more often non-

Hispanic white (55% missing vs. 48% non-missing); more often had their procedure in the first year of our study, 2005 (27% missing vs 2% non-missing); more often had shorter duration of observed diabetes (68% missing vs. 51% non-missing); more often had missing blood pressure and triglyceride measures (39% missing vs 0.5% non-missing); and more often had missing information on smoking status (76% missing vs.2% non-missing).

Missing data were encountered at baseline for race/ethnicity, self-reported smoking status, blood pressure, and triglyceride levels. Table 1 in the main article displays the amount of missing data for each variable, stratified by surgery/control status. Multiple imputation via chained equations using the ICE MI function in Stata was performed to model for missing data. Race/ethnicity and smoking status were imputed using multinomial logistic regression, while uncontrolled blood pressure, LDL cholesterol, and high triglycerides were imputed via logistic regression. eGFR was imputed using a linear regression model. No problems related to perfect prediction or nonconvergence of the models were experienced. The imputation used all variables involved in the four analytic models, including the outcome variables of time-to-event and event status. We used M=10 imputations, with 100 iterations between saved datasets to prevent autocorrelation between imputations. Stata's "mi estimate:" prefix was used to automatically combine the results of the Cox regressions using Rubin's rules.

Sensitivity Analysis for Unmeasured Confounding

Additional sensitivity analysis for the potential effect of unmeasured confounding was performed by the E-value methodology of VanderWeele and Ding (Annals of Internal Medicine, 2017 Aug 15;167(4):268-74. This method estimates the minimum strength of association that would be required between an unmeasured confounder and both receipt of bariatric surgery and risk of incident macrovascular disease to overcome the statistically significant effect observed in a study where residual confounding is a potential problem. The calculation is derived from the relative risk or HR obtained from an adjusted analysis in observational studies.

For the current study, the HR at 5-years following bariatric surgery for macrovascular disease development was 0.60 with a 95% CI of 0.42-0.86 (Table 3 of the main article). The E-value for this point estimate is 2.72 and for the upper confidence interval limit is 1.60. Thus, following the suggested language of VanderWeele and Ding, we found that the observed HR of 0.60 could be explained by an unmeasured confounder that was associated with both receipt of bariatric surgery and risk of incident macrovascular disease by a risk ratio of 2.72 each, above and beyond the measured confounders, but weaker confounding could not do so.

In the current study, it seems implausible to have an unmeasured confounder with this large of an association with both receipt of bariatric surgery and risk of incident macrovascular disease, especially given that it is much larger than that observed for well known risk factors for macrovascular disease such as hypertension, diabetes, and hyperlipidemia (see Table 4 main article). One limitation for how we employed the E-value methodology is that we used a time-varying hazard ratio instead of a single (proportional) hazard ratio.

Software used

Throughout, we used *SAS* version 9.4 (SAS Institute, Cary NC) for data manipulation, *Stata 15.1* for multiple imputation and analysis, and *R* for visualization.

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

eTable 1. Baseline Characteristics of Patients With Type 2 Diabetes Who Underwent Bariatric Surgery Who Were Included in the Study and Those That Were Excluded Due to Missing Data or Inability to Match to Nonsurgical Patients, 2005-2011*

	Patients with Bariatric Surgery	**Excluded due to missing data or no matches**
Number	5,301	990
Age (years), mean (SD)	49.5 (10.0)	47.8 (10.4)
Age Categories, N (%)		
18-29 y	118 (2%)	52 (5%)
30-44 y	1531 (29%)	315 (32%)
45-54 y	1857 (35%)	350 (35%)
55-64 y	1517 (29%)	235 (24%)
65-79 y	278 (5%)	38 (4%)
Female, N (%)	4023 (76%)	745 (75%)
Race/Ethnicity, N (%)		
Hispanic	930 (18%)	112 (11%)
Non-Hispanic black	812 (15%)	116 (12%)
Non-Hispanic white	2534 (48%)	542 (55%)
Other	400 (8%)	45 (5%)
Unknown/Missing	625 (12%)	175 (18%)
Health care site, N (%)		
HealthPartners (HP)	247 (5%)	343 (35%)
Kaiser Permanente Northern California (KPNC)	1290 (24%)	49 (5%)
Kaiser Permanente Southern California (KPSC)	3381 (64%)	461 (47%)
Kaiser Permanente Washington (KPWA)	383 (7%)	137 (14%)
Insurance type, N (%)		
Commercial	4908 (93%)	892 (90%)
Medicaid	144 (3%)	30 (3%)
Medicare	172 (3%)	40 (4%)
Other	77 (1%)	28 (3%)
Year of surgery/index date, N (%)		
2005	87 (2%)	270 (27%)
2006	304 (6%)	208 (21%)
2007	536 (10%)	164 (17%)
2008	769 (15%)	87 (9%)
2009	852 (16%)	84 (8%)
2010	1163 (22%)	105 (11%)
2011	1590 (30%)	72 (7%)

	Patients with Bariatric Surgery	Excluded due to missing data or no matches
Total number of days of health care use in 7-24 months pre-index date, mean (SD)	20.0 (12.7)	6.1 (9.9)
BMI, kg/m^2, mean (SD)	44.7 (6.9)	49.7 (11.3)
BMI Categories, kg/m^2, N (%)		
35.0-39.9	1460 (28%)	38 (4%)
40.0-49.9	2781 (52%)	137 (14%)
50.0+	1060 (20%)	100 (10%)
eGFR (SD), mL/min/1.73 m^2	90.5 (23.0)	91.4 (23.3)
HbA1c %, mean (SD)	7.17 (1.24)	7.35 (1.63)
Observed duration of diabetes in years, mean (SD)	5.64 (4.06)	3.94 (3.26)
0-4 years	2704 (51%)	670 (68%)
5+ years	2597 (49%)	320 (32%)
Use of oral diabetes medication, N (%)	3639 (69%)	644 (65%)
Use of insulin, N (%)	1294 (24%)	247 (25%)
Peripheral arterial disease, N (%)	175 (3%)	41 (4%)
Dyslipidemia, N (%)		
Triglyceride level ≥1.7 mmol/L (≥150 mg/dL)	2483 (47%)	303 (31%)
Missing triglyceride level	51 (1%)	384 (39%)
Dyslipidemia diagnosis*	4405 (83%)	784 (79%)
Use of a statin	2893 (55%)	462 (47%)
Use of other lipid-lowering medications	337 (6%)	80 (8%)
LDL level ≥2.59 mmol/L (≥100 mg/dL)	855 (16%)	207 (21%)
Missing LDL level	3527 (67%)	627 (63%)
Hypertension, N (%)		
Uncontrolled hypertension	422 (8%)	29 (3%)
Missing BP measurement	23 (0%)	718 (73%)
Hypertension diagnosis*	4080 (77%)	748 (76%)
Use of ACE inhibitors or ARBs	2992 (56%)	527 (53%)
Use of other antihypertensive medications	2421 (46%)	466 (47%)
Microvascular disease		
Diabetes with renal manifestations, end stage renal disease, or dialysis*	1256 (24%)	173 (17%)
Diabetic retinopathy*	642 (12%)	102 (10%)
Diabetic neuropathy*	1050 (20%)	176 (18%)
Smoking status*		
Current	469 (9%)	29 (3%)

	Patients with Bariatric Surgery	Excluded due to missing data or no matches
Former	1701 (32%)	81 (8%)
Never	3009 (57%)	129 (13%)
Missing	122 (2%)	751 (76%)

* BMI = Body mass index; eGFR = estimated glomerular filtration rate; LDL = low density lipoprotein; BP = Blood pressure; ACE = angiotensin converting enzyme; ARB = angiotensin receptor blocker; Values represent characteristics at the time of bariatric surgery for surgical patients (or equivalent index date for non-surgical patients) or for the two-year period prior to surgery (as indicated by *). For categorical variables, counts and percentages are presented; for continuous variables, means ± standard deviations are presented.

eTable 2. Reasons for Censoring for Surgical and Nonsurgical Populations

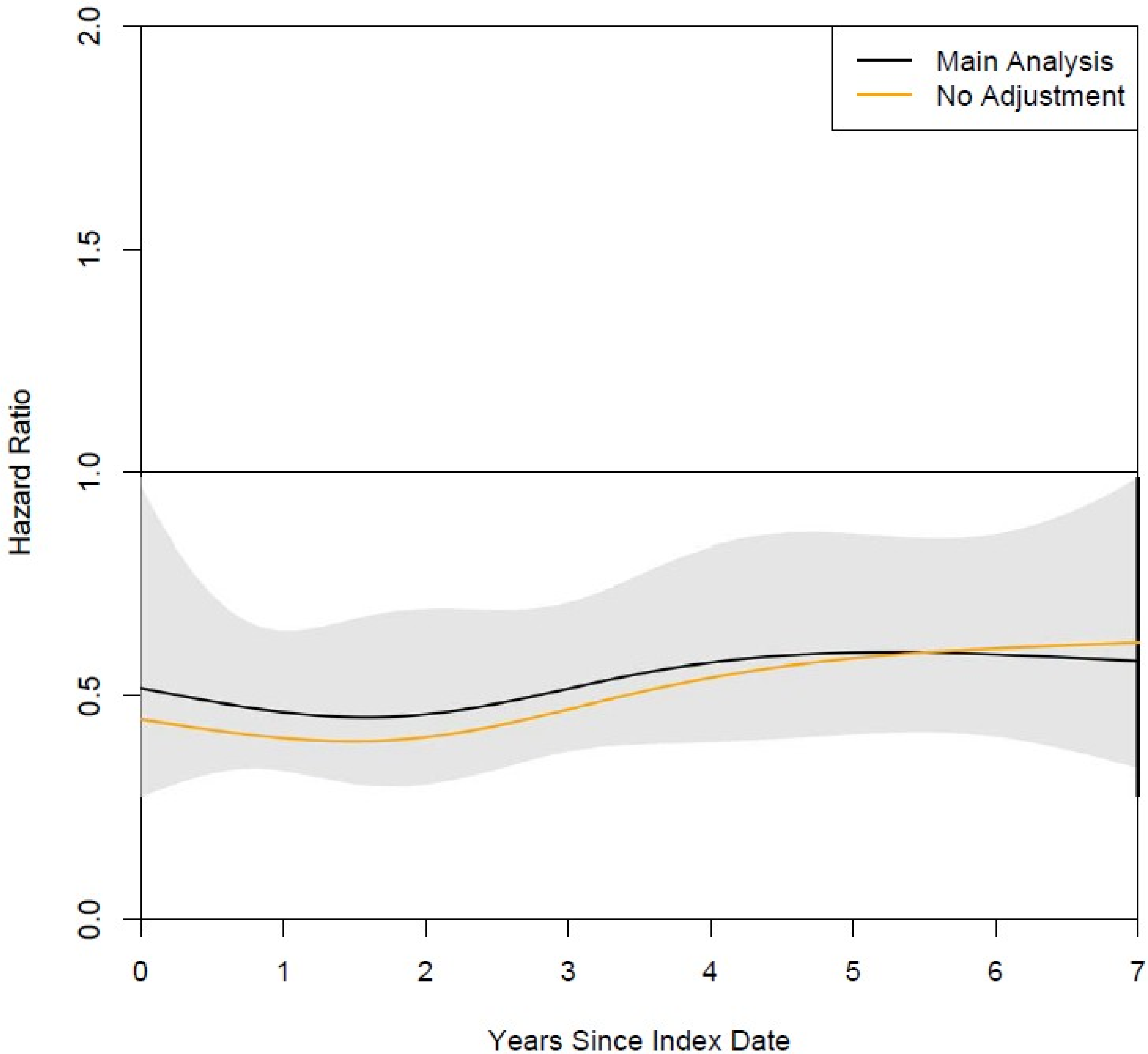
Non-Surgical Population					
	1 year	2 years	3 years	4 years	5 years
Incident Cancer	175 (1%)	325 (2%)	449 (3%)	584 (4%)	687 (5%)
Death	124 (1%)	245 (2%)	357 (2%)	455 (3%)	525 (4%)
Disenrollment	1159 (8%)	2107 (14%)	2818 (19%)	3358 (22%)	3767 (25%)
Study End	0 (0%)	0 (0%)	0 (0%)	825 (6%)	3856 (26%)
Not Censored	13476 (90%)	12257 (82%)	11310 (76%)	9712 (65%)	6099 (41%)
Surgical Population					
	1 year	2 years	3 years	4 years	5 years
Incident Cancer	50 (1%)	90 (2%)	127 (2%)	166 (3%)	191 (4%)
Death	20 (0%)	36 (1%)	42 (1%)	50 (1%)	58 (1%)
Disenrollment	427 (8%)	771 (15%)	1050 (20%)	1286 (24%)	1442 (27%)
Study End	0 (0%)	0 (0%)	0 (0%)	264 (5%)	1332 (25%)
Not Censored	4804 (91%)	4404 (83%)	4082 (77%)	3535 (67%)	2278 (43%)

eTable 3. Cumulative Number of Incident Macrovascular Events and Deaths Among Bariatric Patients and Matched Nonsurgical Patients

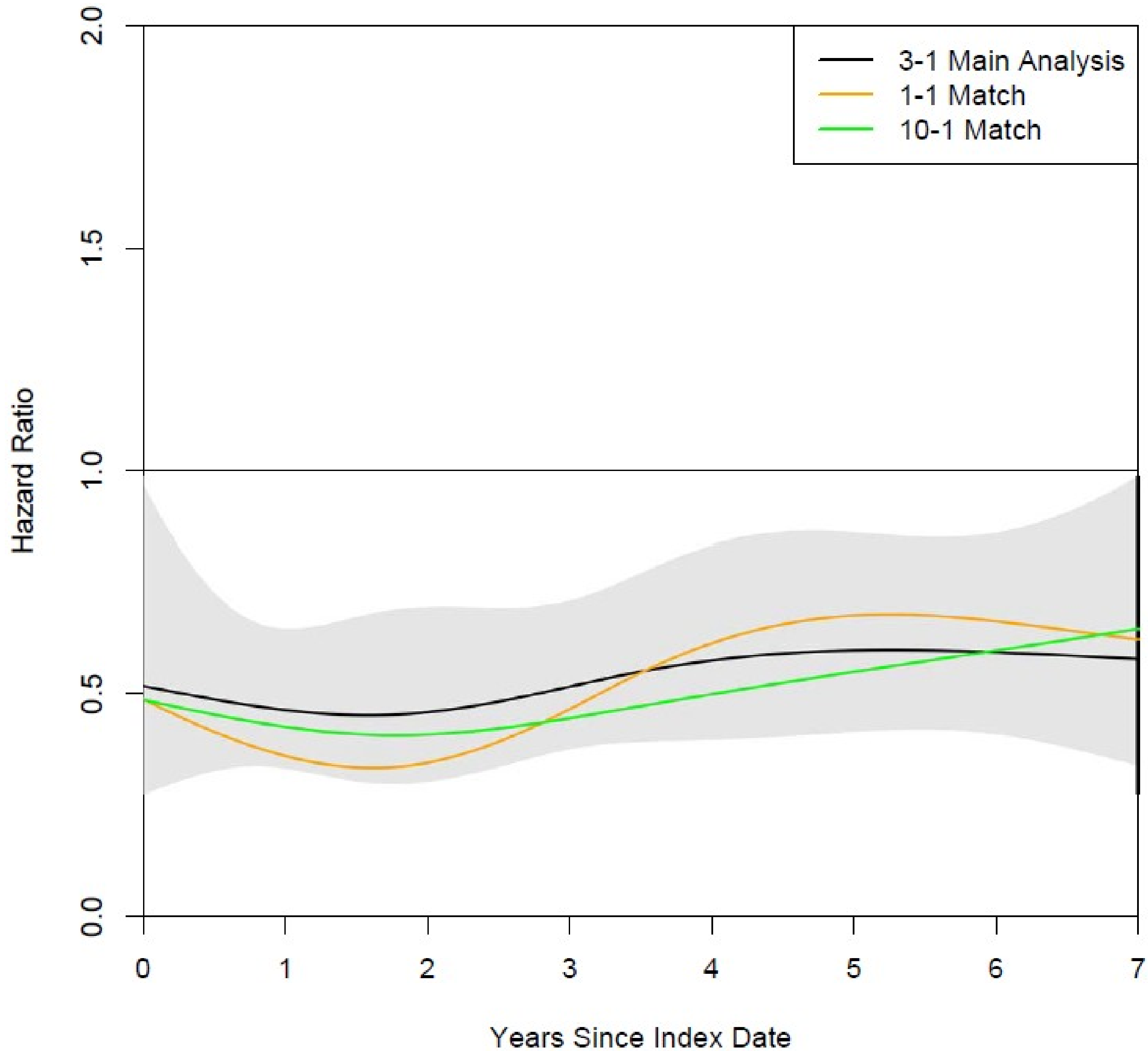
		Number of events over time				
		1 Year	3 Years	5 Years	7 Years	Total
All Macrovascular	Surgical	24	52	85	100	106
	Non-surgical	155	348	502	573	596
Coronary artery disease	Surgical	18	41	65	74	78
	Non-surgical	111	237	332	384	398
Cerebrovascular disease	Surgical	10	16	28	35	37
	Non-surgical	54	131	196	218	227
Death	Surgical	20	41	58	68	70
	Non-surgical	143	387	538	608	622

eFigure. Sensitivity Analyses

eFigure, A. Sensitivity Analyses comparing the time varying hazard ratio for all macrovascular events (composite endpoint) in our main analysis (fully adjusted) versus a matched, unadjusted analysis (no adjustment)



eFigure, B. Sensitivity Analyses comparing the time varying hazard ratio for all macrovascular events (composite endpoint) in our main analysis (3:1 matching) versus alternative 10:1 and 1:1 matching approaches



eFigure, C. Sensitivity Analyses comparing the time varying hazard ratio for all macrovascular events (composite endpoint) in our main analysis (fully adjusted) versus a similar fully-adjusted model that includes all covariates in the Malahanobis distance calculation for our matching process.