Supplementary Materials

MADCO-PC study Inclusion/Exclusion Criteria:

Participants were excluded if they were unable to read or speak English, imprisoned, taking anticoagulants that would preclude lumbar puncture,[1] taking systemic chemotherapy, had a bleeding/clotting disorder, or had a personal or family history of malignant hyperthermia or other medical contraindication to standard anesthetic drugs. Patients were not excluded from study participation for having an anticipated postoperative ICU stay or if it was unclear if they could be extubated within 24 hours after surgery. However, most of the surgical patients underwent routine procedures that took 2-3 hours (ie, elective joint arthroplasties, thyroidectomies, prostatectomies, etc), and which are unlikely to be followed by an ICU stay or postoperative intubation. Nonsurgical controls were excluded from further study participation if they underwent surgery during the study period.

CSF AD Biomarker Procedures

The first 4 surgical participants had preoperative and 24-hour postoperative CSF samples collected via intrathecal catheters (Wiley Spinal; Dallas, TX) placed at the L4-L5 or L5-S1 interspace. We subsequently switched to obtaining all CSF samples via lumbar puncture for the 6 week postoperative samples for these 4 participants, and for all subsequent study patients as described.[2] CSF samples were immediately placed in a pre-chilled 15ml conical tube (VWR; Radnor, PA) on ice. Samples were then separated into 1ml aliquots using low-binding 1000µl pipette tips (Genesee; San Diego, CA), placed in pre-chilled Sarstedt 1.5ml polypropylene microcentrifuge tubes (VWR; Radnor, PA) and stored at -80°C.

CSF sample aliquots were shipped on dry ice to the Alzheimer Disease Neuroimaging Initiative (ADNI) Biomarker Core Laboratory at the University of Pennsylvania for quantification of AD biomarker levels, to avoid inter-laboratory measurement variance.[3] Levels of $A\beta_{1-42}$, t-tau and p-tau-181p (tau phosphorylated at threonine 181) were measured

using the multiplex xMAP Luminex platform (Luminex Corp; Austin, TX) and Innogenetics (INNO-BIA AlzBio3; Ghent, Belgium) immunoassay kit reagents according to the ADNI protocol.[4] A small number of participants had missing CSF samples due to refusal of or inability to perform the lumbar puncture, thus excluding N = 9, 17 and 13 surgical participants and N = 0, 4, and 10 nonsurgical controls from baseline, 24 hours, and 6-week CSF AD biomarker analyses, respectively.

CSF biomarker levels (in pg/ml) were analyzed as continuous measures. Additionally, participants were considered to have brain Aβ pathology (i.e., Aβ+) if their CSF Aβ levels were < 250 pg/ml, and brain tau pathology (i.e. Tau+) if their CSF tau levels were > 93 pg/ml, based on published threshold values.[5,6] The AlzBio3 tau assay has better accuracy than the AlzBio3 p-tau-181p assay for predicting MCI and AD;[5] thus, we used CSF tau levels (rather than p-tau-181p) to define the presence of brain tau pathology.

Since the AlzBio3 and Lumipulse assay platform scales for these biomarkers were based on different standards, we report the baseline, 24-hour and 6 week timepoint data from the AlzBio3 assay, with separately presented Lumipulse 1 year timepoint data.

Cognitive Test Procedures

The following tests were included in the cognitive battery, grouped by cognitive factor domains: Verbal Memory—Randt Short Story Memory Test,[7] Visual Memory—Modified Visual Reproduction Test from the Wechsler Adult Intelligence Scale-Revised (WAIS-R),[8] Executive Function—Digit Symbol from the WAIS-R[8] and the Trail Making Test Parts A and B,[9] and Attention/Concentration—Digit Span from the WAIS-R.[8] Factor analysis of subtest scores from the cognitive battery utilized factor loadings from a prior study[2] of older non-cardiac surgical patients that completed the same cognitive battery, as previously reported.[10] Since the Trail Making Test scale is in the opposite direction from the rest of the tests in the battery, we negatively log transformed the trails scores before including them in the factor analysis.

Cognitive tests were given to study participants by our neurocognitive testing staff, who were trained in test administration by a neuropsychologist.

Mild NCD was defined as a subjective cognitive complaint 6 weeks or 1 year after surgery and a ≥ 1 SD but < 2 SD decrease in cognitive scores in at least 1 of the 4 cognitive domains from before to 1 year after surgery.[11] Major NCD was defined as a subjective cognitive complaint, a ≥ 2 SD decrease in cognitive scores in at least 1 of the 4 cognitive domains and a worsening of IADL scores, at 6 weeks or 1 year after surgery.

Quality of Life Questionnaires

Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D is a 20-item measure of depressive symptomology with an emphasis on depressive mood and affect such as feelings of hopelessness and helplessness with scores ranging from 0 to 60 (higher scores indicate more depressive symptomology). The CES-D has high sensitivity and specificity as a screening tool for major depression in older primary care patients.[12]

Duke Activity Status Index (DASI). The DASI is a 12-item self-report questionnaire measuring physical functional capacity with a score range of 0 to 60. Preoperative DASI scores below 34 have been predictive of 30-day and 1-year postoperative complications among non-cardiac, non-neurologic surgical patients.[13]

The Medical Outcomes Study 36-Item Short Form Health Survey (SF-36).[14,15] The SF-36 is a 36-item self-report health survey measuring eight major health dimensions, including general health perceptions and work activities. General health perceptions (one item) probes the participant's subjective view of their overall health (either mental or physical—not specified in the questions), while work activities (four items) assesses the degree to which the participant's health has interfered with work or other activities.

Instrumental Activities of Daily Living (IADL). This survey measures the participant's ability to perform independent tasks such as cooking, cleaning, and managing finances,

transportation, and medication and instrumental activities, as previously described.[16] IADL impairments or disabilities have been associated with mild cognitive impairment in older adults.[17,18]

Cognitive Difficulties Scale (CDS). The CDS is a 39-item measure of perceived difficulties in memory or other cognitive functions as rated on a 5-point Likert scale.[19]

Social Activity.[20] This scale consists of 8 items that assesses an individual's degree of social interaction with questions such as "How often do you talk on the telephone with friends and relatives?"

Perceived Social Support Scale (PSSS).[21] The PSSS measures an individual's subjective rating of social support using 12 items on a 5-point Likert scale and has been tested among patients undergoing diagnostic coronary angiography.[22]

State-Trait Anxiety Inventory (STAI).[23] The STAI has two components—a 20-item measure of transient state anxiety and 20-item measure of chronic trait anxiety—which has been validated in a geriatric population.[24] This study only included the state anxiety subscale.

Symptom Limitations Scale.[25] Participants are asked to rate the degree to which certain symptoms (angina, shortness of breath, arthritis, back trouble, leg pains, headaches, fatigue, and other) limit daily activities.

Work Status.[26] This single-item question asks for the best description of current working status: working full-time, working part-time, on short-term sick leave, on long-term sick leave, temporarily laid off, homemaker, disabled, unemployed or looking for work, retired, or other (specified). If working full-time or part-time, patients were asked on which date they returned to work after surgery.

Postoperative Delirium Assessment

Postoperative delirium (POD) incidence was measured twice daily through postoperative day 5 or hospital release using the Confusion Assessment Method (CAM).[27] To further increase

sensitivity for delirium detection (since a patient could have tested normal on the CAMs yet had delirium at another point during the day due to the fluctuating nature of delirium), we also used a validated chart review method for detecting delirium.[28] Patients were defined as having delirium if they tested positive on the CAM at any time point during their postoperative hospitalization or if they had a positive chart review for delirium at any point during their postoperative hospitalization.

Enrollment Follow-up Details and Model Sample Sizes

Of the 185 consented participants (137 surgical, 48 nonsurgical) that completed baseline cognitive testing, 120 (80 surgical, 40 nonsurgical) returned at 1-year. After excluding 7 surgical patients missing baseline CSF data for Aβ|tau status classification, 5 nonsurgical controls missing hypertension data, and 1 surgical patient missing DASI and Symptom Limitations scores from the total sample of 120, data remaining from 107 participants (72 surgical, 35 nonsurgical) at baseline and 1-year were included in the multivariable 1-year cognitive change model. Similarly, the non-parametric longitudinal modeling approach for CSF AD biomarker data used all available data at each time point with N=147 (101 surgical, 46 nonsurgical), 135 (93 surgical, 42 nonsurgical), and 133 (97 surgical, 36 nonsurgical) observed CSF AD biomarker datapoints from the surgical patients and nonsurgical controls at baseline, 24 hours, and 6 weeks, respectively. One additional nonsurgical control was missing tau data at 24-hours due to an assay artifact. At 1 year, 48 surgical patients and 32 nonsurgical controls among our analysis cohort underwent lumbar punctures to provide CSF, on which Lumipulse CSF assays were performed.

Power and Sample Size Calculations

Power and sample size calculations were performed based on preliminary data from our MAD-PIA trial[29] that measured perioperative CSF AD biomarkers in patients undergoing ENT and neurosurgical procedures. Based on the mean and variance of CSF tau change from before to 24

hours after surgery in that study,[29] 85 surgical and 42 nonsurgical participants would provide 80% power with $\alpha = 0.05$ to detect a 75% smaller increase in CSF tau levels from baseline to 24-hours in nonsurgical controls versus surgical patients. This would correspond to a baseline to 24-hour CSF tau increase of ≤ 22 for nonsurgical controls, assuming a mean baseline to 24-hour CSF tau increase of 87 pg/ml in the surgical patients (as we observed in a prior study[29]). This sample size would also provide the ability to measure a 0.15 unit difference or greater in cognitive change (i.e. continuous cognitive index or CCI change) between surgical patients vs non-surgical controls from before to 6 weeks or 1 year after surgery. This would be practically meaningful, as a 0.15 unit difference in 1-year cognitive change would represent ~half of the difference in cognitive change units between patients with vs without postoperative cognitive dysfunction observed in our prior work.[30]

Data Provenance and Statistics

All paper copies of tests and questionnaires from MADCO-PC were securely stored on site at Duke University, and each data point was entered into once a secure electronic database (i.e. single data entry). For quality control, we rechecked the accuracy of electronically entered cognitive test scores against paper test scores from a random 10% of participants (N=5 nonsurgical controls, 10 surgical patients) with complete baseline and 6-week cognitive scores. All (100%) of these re-checked data points were correctly entered into the secure electronic database and correctly loaded into our statistical software (SAS Studio 3.8). Nonparametric longitudinal models (a non-parametric version of a repeated measures ANOVA) of CSF AD-related biomarker trajectories were analyzed in R 3.5.0 using nparLD.[31]

Multiple Imputation Strategy

For sensitivity analyses, we performed multiple imputation on missing 1-year cognitive data for all patients with complete baseline and 1-year cognitive data (N=137 surgical patients and 48

nonsurgical controls) using SRCWare software v0.2. Imputed values were based on all known cognitive scores, including MMSE, as well as CSF AD biomarker concentrations, baseline quality of life scores, and patient factors such as years of education, age, APOE4 genotype, hypertension diagnosis, race, and gender. SRCware employs multivariate sequential regression to replace missing data within a plausible range to represent uncertainty in the true value. This yields multiple imputed data sets, which are analyzed in the same manner as the rest of the data. The analyses are then pooled to produce results that properly reflect uncertainty resulting from the missing data points.[32] We created 10 imputation datasets and used standard methods to pool across imputed sets, as was done previously.[2]

Inverse Probability Weighting (IPW)

Inverse probability weighting (IPW) is an approach to handle missing data by weighting the completely observed data by the probability of being non-missing.[33,34] We used a logistic regression model based on observed baseline data among all enrolled subjects to predict the probability of having non-missing outcome data. We assumed that, based on all observed data elements, we can correctly estimate this probability of non-missingness, and that there are representative subjects that did return for those subjects with missing data. Using the model estimated probability of non-missingness, we construct weights as the inverse of the predicted probability. Those weights were then applied to the complete data subjects in our cognitive outcome models to make inference for both the fully observed and the unobserved cohort. This analysis was performed in SAS v 9.4, and the weights were calculated and model diagnostics performed using proc PSMATCH. Weights were generated using all available baseline cognitive scores as well as baseline age, years of education, MMSE score, gender, race, APOE4 genotype, and quality of life scores. Since IPW relies on complete cases for weight calculation, and some participants had missingness in variables like years of education, the final IPW dataset consisted of 48 participants with data lost to follow-up and 117 participants with complete cases.

Tipping Point Approach

This is an approach for data that may be missing not at random. This approach calculates the required shift in the imputed 1-year cognitive decline scores of missing surgical patients that would be necessary to find that surgical patients had significantly greater cognitive decline at 1-year than nonsurgical controls. Then, we determined the plausibility of theoretically observing this calculated shift in cognitive decline scores in the missing surgical patients, based upon the number of standard deviations that this shifted value falls from our observed range of cognitive change scores.

Longitudinal Mixed Model Equations

$$Cog.Func = \beta_0 + \beta_1 Surg + \beta_2 SW + \beta_3 Yr + \beta_4 SurgxSW + \beta_5 SurgxYr + \overline{\beta_6}\boldsymbol{Z}$$

Surg = Group (surgical patients)
SW = 6-week timepoint
Yr = 1-year timepoint
β_6 = Vector of beta's for adjustment variables
$\boldsymbol{Z}$ = Adjustment covariate matrix

Quantities of Interest:

Cognitive Function by Time Point
Baseline: $\beta_0 + \beta_1 Surg + \overline{\beta_6}\boldsymbol{Z}$
6-Week: $\beta_0 + \beta_1 Surg + \beta_2 SW + \beta_4 SurgxSW + \overline{\beta_6}\boldsymbol{Z}$
1-Year: $\beta_0 + \beta_1 Surg + \beta_3 Yr + \beta_5 SurgxYr + \overline{\beta_6}\boldsymbol{Z}$

Cognitive Function by Time Point for Non-Surgical Group
Baseline: $\beta_0 + \overline{\beta_6}\boldsymbol{Z}$
6-Week: $\beta_0 + \beta_2 SW + \overline{\beta_6}\boldsymbol{Z}$
1-Year: $\beta_0 + \beta_3 Yr + \overline{\beta_6}\boldsymbol{Z}$

Cognitive Function by Time Point for Surgical Group
Baseline: $\beta_0 + \beta_1 Surg + \overline{\beta_6}\boldsymbol{Z}$
6-Week: $\beta_0 + \beta_1 Surg + \beta_2 SW + \beta_4 SurgxSW + \overline{\beta_6}\boldsymbol{Z}$
1-Year: $\beta_0 + \beta_1 Surg + \beta_3 Yr + \beta_5 SurgxYr + \overline{\beta_6}\boldsymbol{Z}$

Interest in Difference in Cognitive Change from Baseline to 1-Year
Year – Baseline = $(\beta_0 + \beta_1 Surg + \beta_3 Yr + \beta_5 SurgxYr + \overline{\beta_6}\boldsymbol{Z}) - (\beta_0 + \beta_1 Surg + \overline{\beta_6}\boldsymbol{Z})$

$$= \beta_3 Yr + \beta_5 SurgxYr$$

For nonsurgical controls, this is: $\beta_3 Yr$
For surgical patients, this is: $\beta_3 Yr + \beta_5 SurgxYr$

Hence, the difference in cognitive change from baseline to 1-year between nonsurgical controls and surgical patients is: $\beta_5 SurgxYr$

Longitudinal Mixed Model Results

In a longitudinal mixed model of cognitive change controlling for Aβ|Tau classification, baseline (BL) DASI, BL SF-36 General Health Perception, BL Social Activities, and BL Symptom Limitations scores, and BL hypertension, there was a significant interaction between group and time for CCI change from baseline to 1-year follow-up (β [95% CI]: -0.30 [-0.43, -0.18], $p < 0.001$; Supplemental Table 9). Thus, the nonsurgical controls had greater cognitive decline over this 1-year interval than the surgical patients (Supplemental Figure 1). Baseline Aβ|tau pathology was independently associated with CCI (-0.44 [-0.70, 0.18], $p<0.001$; Supplemental Table 8). There was no evidence of an Aβ|tau classification by time interaction in the longitudinal mixed model; thus, it was removed. There was an effect of baseline SF-36 General Health Perception scores in the model. Therefore, to study whether the imbalance between groups in SF-36 General Health Perceptions scores at baseline might explain part or all of the differential 1-year cognitive change between groups, we examined whether there was an interaction effect between health perception and time for cognition. Patients with worse perceived health (higher scores) at 1 year had lower cognitive function than patients with better perceived health (-0.22 [-0.36, -0.07], $p=0.004$), although the significant effect of group by time at 1 year remained (Supplemental Table 9).

REFERENCES

1. Horlocker TT, Wedel DJ, Rowlingson JC, et al. Regional anesthesia in the patient receiving antithrombotic or thrombolytic therapy: American Society of Regional Anesthesia and Pain Medicine Evidence-Based Guidelines (Third Edition). *Reg Anesth Pain Med.* 2010;35(1):64-101.
2. Berger M, Browndyke JN, Cooter Wright M, et al. Postoperative changes in cognition and cerebrospinal fluid neurodegenerative disease biomarkers. *Ann Clin Transl Neurol.* 2022;9(2):155-170.
3. Trojanowski JQ, Vandeerstichele H, Korecka M, et al. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. *Alzheimers Dement.* 2010;6(3):230-238.
4. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol.* 2011;121(5):597-609.
5. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol.* 2009;65(4):403-413.
6. Shaw LM, Waligorska T, Fields L, et al. Derivation of cutoffs for the Elecsys((R)) amyloid beta (1-42) assay in Alzheimer's disease. *Alzheimers Dement (Amst).* 2018;10:698-705.
7. Randt C, Brown E. Randt memory test. 1983.
8. Wechsler D. *Wechsler adult intelligence scale-revised (WAIS-R).* Psychological Corporation; 1981.
9. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills.* 1958;8(3):271-276.
10. McDonagh DL, Mathew JP, White WD, et al. Cognitive function after major noncardiac surgery, apolipoprotein E4 genotype, and biomarkers of brain injury. *Anesthesiology.* 2010;112(4):852-859.
11. Evered L, Silbert B, Knopman DS, et al. Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. *Br J Anaesth.* 2018;121(5):1005-1012.
12. Lyness JM, Noel TK, Cox C, King DA, Conwell T, Caine ED. Screening for depression in elderly primary care patients. *Arch Intern Med.* 1997;157:449-454.
13. Wijeysundera DN, Beattie WS, Hillis GS, et al. Integration of the Duke Activity Status Index into preoperative risk evaluation: a multicentre prospective cohort study. *Br J Anaesth.* 2020;124(3):261-270.
14. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ.* 1992;305(6846):160-164.
15. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care.* 1992;30(6):473-483.
16. Fillenbaum G. The Duke older Americans resources and services procedures. *Multidiminesional Functional Assessment of Older Adults Lawrence Erlbaum, Hillsdale, NJ.* 1988.
17. Ahn IS, Kim JH, Kim S, et al. Impairment of instrumental activities of daily living in patients with mild cognitive impairment. *Psychiatry Investig.* 2009;6(3):180-184.

18. Rajan KB, Hebert LE, Scherr PA, Mendes de Leon CF, Evans DA. Disability in basic and instrumental activities of daily living is associated with faster rate of decline in cognitive function of older adults. *J Gerontol A Biol Sci Med Sci.* 2013;68(5):624-630.
19. McNair DM, Kahn RJ. Self-assessment of cognitive deficits. *Assessment in geriatric psychopharmacology.* 1983;137:143.
20. Phillips-Bute B, Mathew JP, Blumenthal JA, et al. Association of neurocognitive function and quality of life 1 year after coronary artery bypass graft (CABG) surgery. *Psychosom Med.* 2006;68(3):369-375.
21. Zimet G, Dahlem N, Zimet S, Farley G. The perceived social support scale. . Unpublished manuscript.
22. Blumenthal JA, Burg MM, Barefoot J, Williams RB, Haney T, Zimet G. Social support, type A behavior, and coronary artery disease. *Psychosomatic medicine.* 1987;49(4):331-340.
23. Spielberger CD. State-trait anxiety inventory for adults. 1983.
24. Kvaal K, Ulstein I, Nordhus IH, Engedal K. The Spielberger State-Trait Anxiety Inventory (STAI): the state scale in detecting mental disorders in geriatric patients. *Int J Geriatr Psychiatry.* 2005;20(7):629-634.
25. Axelrad KJ. Locus of control and causal attributions as they relate to expectations for coping with a heart attack. 1982.
26. Newman MF, Grocott HP, Mathew JP, et al. Report of the substudy assessing the impact of neurocognitive function on quality of life 5 years after cardiac surgery. *Stroke.* 2001;32(12):2874-2881.
27. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method: a new method for detection of delirium. *Annals of internal medicine.* 1990;113(12):941-948.
28. Inouye SK, Leo-Summers L, Zhang Y, Bogardus Jr ST, Leslie DL, Agostini JV. A chart-based method for identification of delirium: validation compared with interviewer ratings using the confusion assessment method. *Journal of the American Geriatrics Society.* 2005;53(2):312-318.
29. Berger M, Nadler JW, Friedman A, et al. The Effect of Propofol Versus Isoflurane Anesthesia on Human Cerebrospinal Fluid Markers of Alzheimer's Disease: Results of a Randomized Trial. *J Alzheimers Dis.* 2016;52(4):1299-1310.
30. VanDusen KW, Li YJ, Cai V, et al. Cerebrospinal Fluid Proteome Changes in Older Non-Cardiac Surgical Patients with Postoperative Cognitive Dysfunction. *J Alzheimers Dis.* 2021;80(3):1281-1297.
31. Noguchi K, Gel YR, Brunner E, Konietschke F. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *J Stat Softw.* 2012;50(12):1-23.
32. Yuan Y. Multiple Imputation Using SAS Software. *J Stat Softw.* 2011;45(6):1-25.
33. Perkins NJ, Cole SR, Harel O, et al. Principled Approaches to Missing Data in Epidemiologic Studies. *Am J Epidemiol.* 2018;187(3):568-575.
34. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research.* 2013;22(3):278-295.

Supplemental Table 1. Strata-based matching of surgical patients and nonsurgical controls by gender, age, and education level groups/strata.

Strata Group Number	Gender	Age Groups (Years)	Education Level (Years)	Final Surgical Cohort with 6-Week Follow-up	Planned Nonsurgical Control Recruitment	Final Nonsurgical Control Recruitment
1	**F**	**60-64**	**<12**	0	0	0
2	**F**	**60-64**	**12**	1	1	1
3	**F**	**60-64**	**13-15**	3	1	2
4	**F**	**60-64**	**16+**	8	4	5
5	**F**	**65-69**	**<12**	1	0	0
6	**F**	**65-69**	**12**	2	1	0
7	**F**	**65-69**	**13-15**	3	1	1
8	**F**	**65-69**	**16+**	7	3	4
9	**F**	**70-73**	**<12**	0	0	0
10	**F**	**70-73**	**12**	1	1	0
11	**F**	**70-73**	**13-15**	1	0	0
12	**F**	**70-73**	**16+**	3	1	1
13	**F**	**Over 73**	**<12**	0	0	0
14	**F**	**Over 73**	**12**	4	2	2
15	**F**	**Over 73**	**13-15**	4	2	3
16	**F**	**Over 73**	**16+**	4	2	2
17	**M**	**60-64**	**<12**	0	0	0
18	**M**	**60-64**	**12**	3	2	1
19	**M**	**60-64**	**13-15**	6	3	4
20	**M**	**60-64**	**16+**	8	4	4
21	**M**	**65-69**	**<12**	0	0	0
22	**M**	**65-69**	**12**	7	3	1
23	**M**	**65-69**	**13-15**	4	2	3
24	**M**	**65-69**	**16+**	9	4	4
25	**M**	**70-73**	**<12**	1	0	0
26	**M**	**70-73**	**12**	5	2	1
27	**M**	**70-73**	**13-15**	1	0	0
28	**M**	**70-73**	**16+**	12	6	5
29	**M**	**Over 73**	**<12**	2	1	0
30	**M**	**Over 73**	**12**	1	1	1
31	**M**	**Over 73**	**13-15**	3	1	1
32	**M**	**Over 73**	**16+**	6	3	5
	Totals			110*	51	51

*140 surgical patients completed baseline testing; 110 surgical patients returned for 6-week follow-up and these complete cases were used for matching.

Supplemental Table 2. Intraoperative factors in the surgical cohort (N=133 with available intraoperative data). *Of 140 patients. **aaMAC values represent mean values among patients who received a volatile anesthetic for more than 50% of their case. ***Doses are summarized among patients who received the agent. ****Non-depolarizing paralytic doses are given in rocuronium equivalents (RE = Rocuronium + (6/5)*Atracurium + 6*Vecuronium), which specifies that the potency of 1 mg vecuronium is equivalent to 6 mg rocuronium and that 1 mg of atracurium is equivalent to 6/5 mg rocuronium. *****Oral morphine equivalent dose of intraoperative fentanyl and hydromorphone dosage.

Surgical Features		**N(%) or Median [Q1, Q3]**
Surgical Service	General Surgery	40 (30.1%)
	Gynecology	4 (3.01%)
	Orthopedics	21 (15.8%)
	Otolaryngology	3 (2.26%)
	Plastic Surgery	4 (3.01%)
	Thoracic	16 (12.0%)
	Urology	45 (33.8%)
Surgery Duration (minutes)		145.00 [100, 206]
Received volatile anesthetic		69 (51.9%)
Received regional nerve block*		19 (13.6%)
Received epidural*		2 (1.43%)
Mean end tidal aaMAC**		0.75 [0.71, 0.86]
aaMAC Hours**		1.60 [1.28, 2.83]
Available BIS data		85 (63.9%)
Mean BIS value		46 [42, 52]
Intraoperative Medications* **		
Received Acetaminophen		31 (23.3%)
Acetaminophen (mg)		1000 [1000, 1000]
Received Midazolam		63 (47.4%)
Midazolam (mg)		2.00 [2.00, 2.00]
Received Phenylephrine		101 (75.9%)
Phenylephrine (µg)		1191 [360, 2817]
Received Propofol		133 (100%)
Propofol (mg)		330.00 [150, 1273]
Received Succinylcholine		34 (25.6%)
Succinylcholine (mg)		100 [100, 120]
Received Ketamine		26 (19.5%)
Ketamine (mg)		50 [30, 75]
Received Ketorolac		12 (9.0%)
Ketorolac (mg)		15 [15, 15]
Received Dexmedetomidine		15 (11.3%)
Dexmedetomidine (µg)		12 [8, 20]
Received Remifentanil		15 (11.3%)
Remifentanil (µg)		1921 [1156, 2168]
Received Non-depolarizing Paralytics		104 (78.2%)
Non-depolarizing Paralytics (mg)****		65 [50, 100]
Received opioids (ie Oral Morphine Equivalents)		132 (99.2%)
Oral Morphine Equivalents (mg)*****		25 [20, 35]

Supplemental Table 3. Nonparametric longitudinal models of group (surgical patients vs nonsurgical controls), time (baseline, 24-hours, and 6-weeks after surgery), and a group by time interaction for each of 5 CSF AD biomarkers with and without Bonferroni correction for the 5 models.

CSF AD Biomarker Model	Variable	Nonparametric ANOVA-Type Statistic	P-Value	Adjusted P-Value
Aβ	Group	0.59	0.442	>0.999
	Time	1.00	0.366	>0.999
	Group*Time	2.32	0.098	0.490
Tau	Group	3.72	0.054	0.270
	Time	1.14	0.318	>0.999
	Group*Time	2.48	0.087	0.435
P-Tau	Group	3.92	0.048	0.240
	Time	1.50	0.224	>0.999
	Group*Time	2.45	0.091	0.455
Tau/Aβ	Group	2.53	0.112	0.560
	Time	2.10	0.129	0.645
	Group*Time	1.34	0.260	>0.999
P-Tau/Aβ	Group	2.60	0.107	0.535
	Time	1.35	0.258	>0.999
	Group*Time	0.76	0.461	>0.999

Supplemental Table 4. Mean (SD) or median [Q1, Q3] participant scores by cognitive test.

	Surgical Patients (N=137)			Nonsurgical Controls (N=48)		
Variable	**Baseline**	**6 Weeks**	**1 Year**	**Baseline**	**6 Weeks**	**1 Year**
Digit Symbol	42.66 (11.85)	47.00 (10.89)	47.79 (11.17)	44.71 (10.66)	48.72 (10.34)	47.98 (11.50)
Wechsler Reproduction Delay	5.47 (2.94)	5.97 (2.78)	6.46 (2.86)	5.85 (2.74)	5.69 (3.01)	5.58 (3.01)
Wechsler Reproduction Immediate	6.15 (2.81)	7.01 (2.57)	6.95 (2.66)	6.81 (2.39)	7.17 (2.68)	6.28 (2.71)
Delayed RANDT Gist	6.50 (1.90)	6.34 (2.09)	5.73 (2.13)	6.65 (2.42)	6.02 (2.28)	5.28 (2.42)
Delayed RANDT Verbatim	8.76 (3.33)	8.16 (3.36)	7.00 (3.56)	9.15 (4.08)	8.37 (3.78)	6.53 (3.64)
Immediate RANDT Gist	7.11 (1.57)	6.80 (1.70)	6.34 (1.92)	7.58 (1.67)	6.65 (1.85)	5.85 (2.03)
Immediate RANDT Verbatim	10.49 (3.07)	10.00 (3.27)	8.75 (3.45)	11.50 (3.54)	10.26 (3.52)	7.80 (3.29)
Digit Span Repeat Backwards	6.12 (2.28)	5.85 (2.19)	6.15 (2.10)	6.94 (2.49)	6.48 (2.59)	6.35 (2.37)
Digit Span Repeat Forwards	7.50 (2.02)	7.74 (2.06)	8.08 (2.12)	7.96 (2.35)	7.96 (2.30)	8.80 (1.83)
Trails Making Test Part B*	93 [65, 146]	71 [53, 99]	73 [50.5, 101]	93.5 [65, 141.5]	66 [52, 90]	68.5 [55, 104.5]
MMSE	28.15 (1.78)	28.34 (1.60)	28.55 (1.57)	28.38 (1.55)	28.50 (1.72)	28.28 (2.34)

N=107 surgical patients and 46 nonsurgical controls for all 6-week cognitive tests and 80 surgical patients and 40 nonsurgical controls for all 1-year cognitive tests, unless stated otherwise. See Supplemental Materials for additional details on cognitive test procedures. *Raw time to completion trails scores.

Supplemental Table 5. Multivariable linear regression model for CCI change from baseline to 1-year after surgery with IPW. The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension.

Factor	Beta (95% CI)	P-Value
Baseline Cognition	-0.03 (-0.15, 0.08)	0.532
Nonsurgical Controls	-0.33 (-0.47, -0.18)	<0.001
Aβ|Tau Pathology	-0.02 (-0.18, 0.13)	0.754
Hypertension	0.04 (-0.09, 0.17)	0.526
DASI	0.00 (-0.00, 0.01)	0.390
SF-36 General Health Perceptions	-0.11 (-0.19, -0.02)	0.014
Social Activities	-0.01 (-0.02, 0.01)	0.478
Symptom Limitations	0.01 (-0.02, 0.04)	0.403

Supplemental Table 6. Multivariable linear regression model for CCI change from baseline to 1-year after surgery with imputed baseline data for all variables listed below (N=137 surgical patients, 48 nonsurgical controls). The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension.

Factor	Beta (95% CI)	P-Value
Baseline Cognition	-0.21 (-0.36, -0.06)	0.010
Nonsurgical Controls	-0.16 (-0.32, 0.01)	0.071
Aβ|Tau Pathology	-0.16 (-0.33, 0.01)	0.063
Hypertension	0.06 (-0.08, 0.21)	0.381
DASI	0.00 (-0.00, 0.01)	0.582
SF-36 General Health Perceptions	-0.05 (-0.17, 0.07)	0.408
Social Activities	-0.01 (-0.03, 0.01)	0.552
Symptom Limitations	-0.01 (-0.03, 0.02)	0.612

Supplemental Table 7. Multivariable model of 1-year postoperative continuous cognitive index change with imputed data (for N=137 surgical patients, 48 nonsurgical controls). 10 of the surgical patients (the ones who were lost to follow-up due to death, health issues, or institutionalization) were given a worst-possible 1-year cognitive index score of -3.412 (ie the cognitive index score that would result from the lowest/worst possible score on each individual cognitive test). The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension.

Factor	Beta (95% CI)	P-Value
Baseline Cognition	-0.22 (-0.44, -0.01)	0.038
Nonsurgical Controls	0.03 (-0.27, 0.33)	0.843
Aβ|Tau Pathology	-0.21 (-0.55, 0.13)	0.219
Hypertension	0.19 (-0.08, 0.45)	0.165
Duke Activity Status Index Score	0.00 (-0.00, 0.01)	0.317
SF-36 General Health Perceptions	-0.09 (-0.25, 0.06)	0.247
Social Activities	-0.01 (-0.04, 0.02)	0.444
Symptom Limitations	-0.02 (-0.06, 0.02)	0.330

Supplemental Table 8. Multivariable linear regression model for continuous cognitive index change from baseline to 1-year after surgery with observed baseline data for all variables listed below (N=72 surgical patients, 35 nonsurgical controls). The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension.

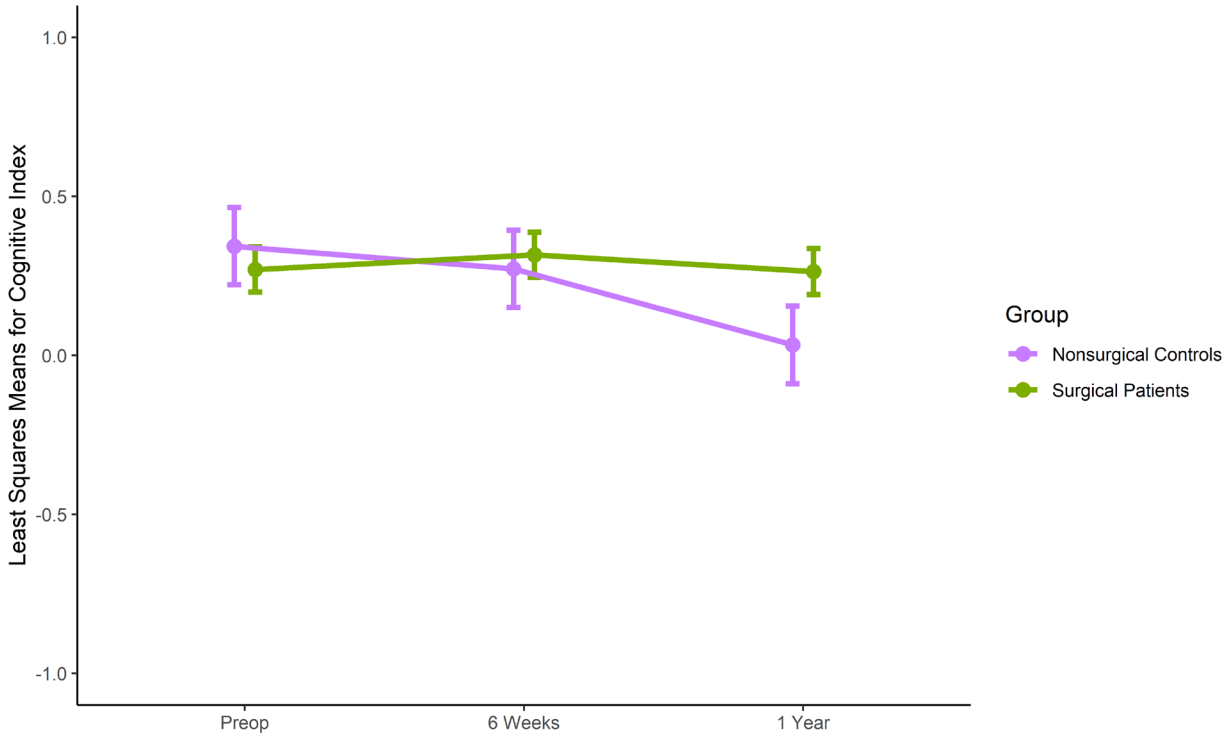
Factor	Beta (95% CI)	P-Value
Baseline Attention/Concentration Score	-0.05 (-0.13, 0.03)	0.189
Nonsurgical Controls	-0.30 (-0.44, -0.16)	<.001
Aβ|Tau Pathology	-0.02 (-0.16, 0.13)	0.833
Hypertension	0.01 (-0.12, 0.14)	0.855
DASI	0.00 (-0.00, 0.01)	0.350
SF-36 General Health Perceptions	-0.11 (-0.19, -0.03)	0.009
Social Activities	-0.00 (-0.02, 0.01)	0.594
Symptom Limitations	0.02 (-0.01, 0.04)	0.190

Supplemental Table 9. Multivariable model of cross-sectional, 1-year postoperative continuous cognitive index values with observed baseline data for all variables listed below (N=72 surgical patients, 35 nonsurgical controls). The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension.

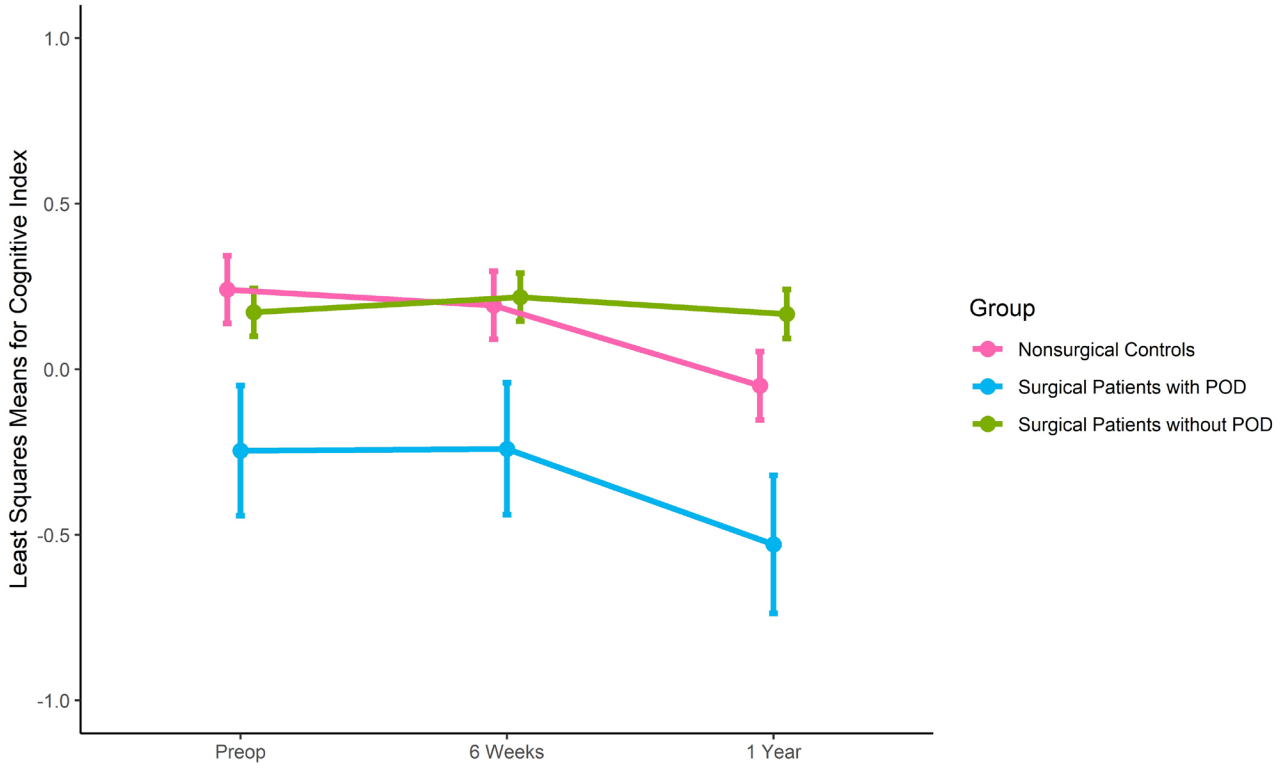
Factor	Beta (95% CI)	P-Value
Baseline Cognition	0.97 (0.86, 1.07)	<.001
Nonsurgical Controls	-0.31 (-0.45, -0.17)	<.001
Aβ|Tau Pathology	-0.02 (-0.17, 0.13)	0.778
Hypertension	0.02 (-0.11, 0.15)	0.735
Duke Activity Status Index Score	0.00 (-0.00, 0.01)	0.249
SF-36 General Health Perceptions	-0.10 (-0.18, -0.02)	0.015
Social Activities	-0.00 (-0.02, 0.01)	0.625
Symptom Limitations	0.02 (-0.01, 0.04)	0.178

Supplemental Table 10. Multivariable mixed model of continuous cognitive index at baseline (BL), 6 weeks, and 1 year with observed data. The reference groups for categorical variables were as follows: surgical patient group, baseline time, Aβ-|tau- classification status, and no hypertension. [*Note:* Beta coefficients for categorical variables (e.g., Aβ|Tau classification) represent differences in cognitive scores between one level of the variable (e.g., individuals with Aβ|Tau pathology) and the reference level (e.g., individuals without Aβ|Tau pathology) while controlling for additional model variables. Group*time interactions represent the difference between groups (nonsurgical vs surgical) in cognitive change from before to 6 weeks or 1-year after surgery (e.g., the beta coefficient of -0.30 for group*time at 1 year indicates that nonsurgical controls had a more negative 1-year cognitive trajectory compared to surgical patients.)]

Factor	Beta (95% CI)	P-Value
Group (nonsurgical vs surgical)	0.06 (-0.20, 0.33)	0.645
Time (6 Weeks)	0.09 (-0.06, 0.25)	0.232
Time (1 Year)	0.24 (0.07, 0.41)	0.006
Interaction of Group*Time (6 Weeks)	-0.12 (-0.23, -0.00)	0.043
Interaction of Group*Time (1 Year)	-0.30 (-0.43, -0.18)	<0.001
BL Aβ|Tau Pathology	-0.44 (-0.70, -0.18)	0.001
BL Hypertension	-0.07 (-0.30, 0.17)	0.574
BL DASI	0.00 (-0.01, 0.01)	0.969
BL SF-36 General Health Perception	-0.12 (-0.26, 0.03)	0.114
BL Social Activities	0.01 (-0.02, 0.04)	0.487
BL Symptom Limitations	-0.02 (-0.06, 0.02)	0.267
Interaction of BL General Health Perception*Time (6 Weeks)	-0.02 (-0.07, 0.03)	0.487
Interaction of BL General Health Perception*Time (1 Year)	-0.10 (-0.16, -0.04)	0.001



Supplemental Figure 1. Plot of the group by time interaction from the mixed model of overall cognitive index (Supplemental Table 9). Nonsurgical controls had a larger decrease in cognitive index from baseline to 1 year compared to surgical patients. The plotted values are adjusted for a mean DASI score of 27.5, mean SF-36 General Health Perceptions score of 2.49, Social Activities score of 15.89, Symptom Limitations score of 12.79, Aβ-|Tau- classification, and no hypertension; error bars represent standard errors after covariate adjustment in the mixed model.



Supplemental Figure 2. Plot of the group by time interaction from the mixed model of overall cognitive index scores among nonsurgical controls and surgical patients with vs without POD. Error bars represent standard errors after covariate adjustment in the mixed model. For nonsurgical controls, N = 48 at baseline, N = 46 at 6 weeks, and N = 40 at 1 year. For surgical patients with POD, N = 13 at baseline, N = 11 at 6 weeks, and N = 7 at 1 year. For surgical patients without POD, N = 96 at baseline, N = 96 at 6 weeks, and N = 73 at 1 year.