



DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU



HARVARD LIBRARY
Office for Scholarly Communication

Delirium Severity Post-Surgery and its Relationship with Long- Term Cognitive Decline in a Cohort of Patients without Dementia

The Harvard community has made this
article openly available. [Please share](#) how
this access benefits you. Your story matters

Citation	Vasunilashorn, Sarinnapha M., Tamara G. Fong, Asha Albuquerque, Edward R. Marcantonio, Eva M. Schmitt, Douglas Tommet, Yun Gou, Thomas G. Trivison, Richard N. Jones, and Sharon K. Inouye. 2017. "Delirium Severity Post-Surgery and Its Relationship with Long-Term Cognitive Decline in a Cohort of Patients Without Dementia." Edited by Miles Berger. <i>Journal of Alzheimer's Disease</i> 61 (1) (November 28): 347–358. doi:10.3233/jad-170288.
Published Version	doi:10.3233/JAD-170288
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:34553726
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

**Delirium Severity Post-Surgery and its Relationship with Long-Term Cognitive Decline in
a Cohort of Patients without Dementia**

Sarinnapha M. Vasunilashorn, PhD,^{a,b*} Tamara G. Fong, MD, PhD,^{b,c,d*}

Asha Albuquerque, BA,^d Edward R. Marcantonio, MD, SM,^{a,b,e} Eva M. Schmitt, PhD,^d

Douglas Tommet, MS,^f Yun Gou, MA,^d Thomas G. Trivison, PhD,^{b,d,g}

Richard N. Jones, ScD,^{f,**} Sharon K. Inouye, MD, MPH,^{b,d,e,**}

*Co-first authors, **Co-senior authors

^aDivision of General Medicine and Primary Care, Department of Medicine, Beth Israel
Deaconess Medical Center

^bHarvard Medical School

^cDepartment of Neurology, Beth Israel Deaconess Medical Center

^dAging Brain Center, Institute for Aging Research, Hebrew SeniorLife

^eDivision of Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center

^fDepartments of Psychiatry and Human Behavior and Neurology, Warren Alpert Medical
School, Brown University

^gResearch Program on Men's Health, Aging, and Metabolism, Brigham and Women's Hospital

Corresponding author:

Sarinnapha M. Vasunilashorn, PhD

Beth Israel Deaconess Medical Center

General Med/CO-1309 – 2nd Fl

330 Brookline Ave

Boston, MA 02215

Email: svasunil@bidmc.harvard.edu

Phone: 617-754-1417

Fax: 617-754-1440

Abstract word count: 250 (250 max)

Manuscript word count: 3,535 (10,000 max)

Tables: 3, Figures: 2

ABSTRACT

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Background: Delirium has been associated with more rapid cognitive decline. However, it is unknown whether increased delirium severity is associated with a higher rate of long-term cognitive decline.

Objective: To evaluate delirium severity and the presence and rate of cognitive decline over 36 months following surgery.

Methods: We examined patients from the Successful Aging after Elective Surgery Study, who were age ≥ 70 years undergoing major elective surgery (N=560). Delirium severity was determined by the peak Confusion Assessment Method-Severity (CAM-S) score for each patient's hospitalization and grouped based on the sample distribution: scores of 0-2, 3-7, and 8-19. A neuropsychological composite, General Cognitive Performance (GCP), and proxy-reported Informant Questionnaire for Cognitive Decline (IQCODE) were used to examine cognitive outcomes following surgery at 0, 1, 2 months, and every 6 months for up to 3 years.

Results: No significant cognitive decline was observed for patients with peak CAM-S scores 0-2 (-0.17 GCP units/year, 95% confidence interval [CI] -0.35, 0.01). GCP scores decreased significantly in the group with peak CAM-S scores 3-7 (-0.30 GCP units/year, 95% CI -0.51, -0.09), and decreased almost three times faster in the highest delirium severity group (peak CAM-S scores 8-19; -0.82 GCP units/year, 95% CI -1.28, -0.37). A similar association was found for delirium severity and the proportion of patients who developed IQCODE impairment over time.

Conclusion: Patients with the highest delirium severity experienced the greatest rate of cognitive decline, which exceeds the rate previously observed for patients with dementia, on serial

61 neuropsychological testing administered over 3 years, with a dose-response relationship
62 between delirium severity and long-term cognitive decline.

63 **Key words:** delirium, cognition, dementia, aged

64 **INTRODUCTION**

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

Delirium is a common and serious problem for hospitalized older persons, associated with prolonged hospital stays, higher hospital costs, increased functional decline, higher rates of institutionalization, and greater mortality [1, 2]. There is growing evidence that delirium is associated with a subsequent course of more rapid cognitive decline [3]. Among patients undergoing cardiac surgery, delirium is associated with a significant decline in cognitive ability, with a trajectory characterized by an initial decline and prolonged impairment [4]. Moreover, patients with Alzheimer’s disease (AD) have a 3-fold increase in the rate of cognitive decline following delirium, compared with those without delirium [5, 6]. In patients without dementia at baseline, those who experienced delirium demonstrated a 4.3-fold greater decline in long-term cognitive performance than the effect of a year of cognitive aging [7]. Although this study [7] and others [8-10] demonstrate that incident delirium is associated with long-term cognitive decline, the critical next step to advance understanding of this relationship is to evaluate whether the severity of delirium is associated with the pace of long-term cognitive decline. This would prove useful for monitoring delirium clinically and for providing a quantifiable dose-response measure for intervention trials seeking to prevent or forestall the long-term cognitive decline associated with delirium.

We have previously shown that delirium severity, as measured by the Confusion Assessment Method-Severity (CAM-S) score [11], demonstrated strong predictive validity for important short-term clinical outcomes associated with delirium, including hospital length of stay, healthcare costs, death, institutionalization, and functional decline [11]. Thus, the Aim of this study was to evaluate whether the severity of delirium was associated with the presence and degree of cognitive decline up to 36 months post-surgery in patients who are free of dementia at

88 baseline. We hypothesized that there would be a graded relationship, with increasing severity of
89 delirium associated with increasing degrees of long-term cognitive decline.

90

91 **MATERIALS AND METHODS**

92

93 Study Population

94 The Successful Aging after Elective Surgery (SAGES) Study is an ongoing prospective
95 cohort study of older adults undergoing major elective non-cardiac surgery. The study design and
96 methods have been previously described [12]. Briefly, eligible participants were age ≥ 70 years,
97 English speaking, scheduled for elective surgery at one of two Harvard-affiliated academic
98 medical centers with an anticipated length of stay ≥ 3 days. Eligible surgical procedures were:
99 total hip or knee replacement, lumbar, cervical, or sacral laminectomy, lower extremity arterial
100 bypass surgery, open abdominal aortic aneurysm repair, and colectomy. Exclusion criteria
101 included evidence of dementia, delirium, hospitalization within 3 months, terminal condition,
102 legal blindness, severe deafness, history of schizophrenia or psychosis, and history of alcohol
103 abuse or withdrawal. A total of 566 patients were eligible and enrolled between June 18, 2010
104 and August 8, 2013. Six patients were subsequently excluded for possible dementia after
105 neuropsychological testing and clinical adjudication (final sample=560; see STROBE diagram
106 and follow-up success rates in the Appendix). This study is in compliance with guidelines on
107 ethical principles for medical research involving human subjects. Written informed consent for
108 study participation was obtained from all participants according to procedures approved by the
109 institutional review boards of Beth Israel Deaconess Medical Center and Brigham and Women's
110 Hospital, the two study hospitals, and Hebrew SeniorLife, the study coordinating center, all
111 located in Boston, Massachusetts.

112

113 Study Procedures

114 Trained research assistants conducted a 90-minute baseline interview in participants’
115 homes about 2 weeks prior to the index surgery [12, 13]. Following surgery, daily interviews
116 were conducted to assess for delirium. After discharge, home-based interviews were conducted
117 by a separate group of trained research assistants (blinded to delirium status) at 1, 2, 6, and every
118 six months up to 36 months. Interviews included assessments of delirium, cognitive and physical
119 function, described below. Medical records were reviewed for the index hospitalization and
120 readmissions.

121

122 Main Study Measures

123 Delirium. The Confusion Assessment Method (CAM) [14] was used to identify delirium
124 at all time points. The CAM provides a standardized method for identification of delirium, with a
125 sensitivity of 94% (95% confidence interval (CI) 91%-97%), specificity of 89% (95% CI 85%-
126 94%), and inter-rater reliability of 0.70-1.00 [15]. All interviewers underwent training and
127 standardization, and inter-rater reliability was determined in 71-paired observations (weighted
128 kappa=0.92) [14]. Delirium was defined as either a positive rating by CAM or by a validated
129 chart review method [16, 17], used to maximize sensitivity.

130 Delirium Severity. The 10-item CAM-S long-form was used to measure delirium severity
131 [11]. Each symptom was rated 0 to 2, except acute onset or fluctuation, which is rated 0 or 1
132 [11], yielding a summary score from 0 to 19 (19=most severe). Because individual patients had
133 multiple CAM ratings during hospitalization, we utilized the highest CAM-S score (peak CAM-
134 S) across all hospital days for each patient to capture the severity of the delirium episode. Peak
135 CAM-S scores (range 0-19) were divided into three groups. Since a minimum of 3 features is
136 required for CAM delirium, the lowest grouping included peak CAM-S scores of 0-2,

137 representing the group without CAM-defined delirium. While the majority of patients without
138 delirium had a score of 0-2, some patients without delirium received higher scores based on non-
139 specific delirium features (e.g., disorientation, memory impairment, psychomotor agitation),
140 which can be present in conditions unrelated to delirium. Next, the group with delirium (N=134)
141 was divided into two groups based on the median peak CAM-S score. These steps allowed
142 delirium patients to be spread across a range of sub-groups rather than clustering only in the
143 highest group, an approach that is preferred when the sample is imbalanced across the
144 distribution [16]. Thus, a single, median-based cutpoint was applied to our patients with SAGES
145 delirium (N=134) (Table 1), resulting in two delirium groups with CAM-S scores of: 1) 3-7 (N=
146 67), and 2) 8-19 (N= 66). These cutpoints were then applied across the entire SAGES cohort.

147 Cognitive Outcome Measures: General Cognitive Performance (GCP) and Informant
148 Questionnaire on Cognitive Decline in the Elderly (IQCODE). A neuropsychological test
149 battery, conducted at baseline and each follow-up, included the Visual Search and Attention Test
150 (VSAT) [20], Hopkins Verbal Learning Test-Revised (HVLTR) [21], Digit Span Forward and
151 Backward [22], Category Fluency (animal naming) [23] Phonemic F-A-S Fluency Tasks [23],
152 Boston Naming Test (BNT) [24], Repeatable Battery for the Assessment of Neuropsychological
153 Status (RBANS) Digit Symbol Substitution Test, Trail-Making Tests (Trails) A and B, and
154 intersecting pentagons from the 3MS [25]. We created a weighted composite summary measure,
155 the GCP score following standard procedures (see [26] for a detailed description). We assessed
156 its reliability and validity and calibrated the GCP score to a nationally representative sample of
157 adults age ≥ 70 years [27] to yield a mean score=50 and standard deviation=10 [25] to improve
158 our ability to make meaningful comparisons to other study populations. The GCP is sensitive to
159 longitudinal change with minimal floor and ceiling effects [26, 28-30].

160 To account for practice effects, GCP scores were adjusted with a correction factor
161 derived from a control sample of comparable non-surgical patients (N=119) from a primary care
162 clinic, who were administered the identical tests on the same schedule (Appendix). Using an
163 accepted approach [31-33], the mean performance of the control sample at each time point was
164 used to center the observed scores in the surgical sample at matching time points. This control
165 group was used only to correct for retest (learning) effects.

166 We used IQCODE [34] as a proxy-reported measure of decline in current abilities for
167 daily cognitive tasks (range 1-5). IQCODE ≥ 3.2 was used to indicate impairment [34].

168 Death and Nursing Home Placement. We examined death or nursing home placement,
169 obtained from patient/proxy interviews and chart review, as a composite outcome between 6-36
170 months follow-up. This timeframe was chosen to indicate long-term outcomes, minimizing acute
171 effects of surgery, hospitalization, or rehabilitation.

172 Other Study Variables. The baseline interview assessed sex, race, ethnicity, education,
173 marital status, living situation, 15-item Geriatric Depression Scale (GDS) [35], Modified Mini-
174 Mental State (3MS) [25], Activities of Daily Living Scale (ADLs) [36], Instrumental Activities
175 of Daily Living Scale (IADLs) [37], and Short Form-12 Health Survey (SF-12) [38]. Age,
176 surgical type, and Charlson comorbidity score [39] were determined from chart review [38].

177

178 Statistical Analyses

179 The overall analytic approaches used general linear mixed effects regression models for
180 the trajectories of GCP score over time. Logistic regression was used for analysis of IQCODE
181 impairment and nursing home placement or death. For GCP, the model included control for
182 delirium severity group, with random effects for baseline GCP level, fixed effects at the 1 and 2
183 month assessments to capture acute decline and recovery, and random effects for linear change

184 after the 2-month follow-up. The delirium severity group variable and the acute decline,
185 recovery, and linear change were regressed on baseline GCP to capture differential effects by
186 baseline status. Therefore, delirium severity group was treated as both an intermediate outcome
187 (dependent upon baseline GCP and covariates) and as a predictor of model parameters capturing
188 GCP change following baseline. Change over time was modeled using a three-part piecewise
189 linear model to describe the longitudinal pattern, including an immediate decline following pre-
190 operative baseline to month 1 (acute decline), recovery from month 1 to 2 following the acute
191 decline (recovery) and long-term trajectory from month 2 to 36 months (long-term trajectory)
192 (Appendix). All models adjusted for baseline covariates, including age, gender, non-white race,
193 education, Charlson score, GDS score, IADL impairment, surgery type, and IQCODE. Analyses
194 were conducted with Mplus (Version 7.4, Muthén & Muthén, Los Angeles, CA).

195 For IQCODE, we used a mixed effects generalized linear model with IQCODE
196 impairment as a repeated outcome at all timepoints. A random effect for the linear slope captured
197 variability in the change over time. For death or nursing home placement, logistic regression was
198 used to model the probability of a participant having the composite of either outcome occurring
199 between months 6–36. Delirium severity was entered as a series of categorical indicators. An
200 interaction between time and delirium severity group captured the differences in linear change
201 over time by severity group. For the death or nursing home analyses with IQCODE, the adjusted
202 models controlled for age, gender, non-white race, education, Charlson score, and surgery type.
203 Baseline IADL and IQCODE were not controlled due to collinearity. Analyses were conducted
204 with Stata software (Version 14.1, Stata Corp, College Station, TX). In analyzing this
205 longitudinal data, our approach to handling data missing at random (MAR) aligns with
206 recommendations by the National Research Council [55].

207 Sensitivity analyses were completed to: (1) assess the extent to which our findings were
208 robust to extreme assumptions regarding cognitive outcomes of persons who left the cohort early
209 due to drop-out, death, or institutionalization (Appendix), and (2) assess the relationship between
210 long-term cognitive decline and sum of all CAM-S scores (an alternate measure of delirium
211 severity that combines both intensity and duration of the delirium episode) [18] (Appendix).

212 **RESULTS**

213 Table 1 reports baseline characteristics overall and stratified by delirium severity group.
214
215 The mean age was 76.7 years, and 58% were women. Delirium occurred in 24%. Forty-four
216 percent had a peak CAM-S score of 0-2; 44% with peak scores of 3-7; and 12% with peak scores
217 of 8-19. Patients with the most severe delirium (peak CAM-S 8-19) were older, had greater
218 impairment on the Charlson, and lower GCP, 3MS, and GDS (all $p < 0.05$). The Spearman rank
219 correlation coefficients indicating the correlation of each variable with the peak CAM-S score
220 are all trivial to moderate in size.

221 The median duration of follow-up for this ongoing cohort was 36 months (interquartile
222 range [IQR] 24-37). Deaths occurred in 7% of patients after a median follow-up of 19 months
223 (IQR 12-26). An additional 27 (5%) participants withdrew from follow-up (i.e., drop-outs) after
224 a median of 5 months (IQR 3-12). Rates of death or drop-out differed between the CAM-S
225 groups, and increased with CAM-S severity level (8%, 12%, and 22% respectively, $p = 0.01$) A
226 total of 496 (89%) eligible participants completed all planned study visits, with a range of 1-9
227 visits per participant. Since this is an ongoing study, the number of visits completed per
228 participant varies according to how long they have been enrolled in the study.

229 We examined cognitive performance by GCP up to 36 months post-surgery (Table 2) by
230 delirium severity. For all groups, GCP scores declined acutely at one month, returned to baseline
231 or above by two months, then remained stable to 3 years, except for the highest severity group

233 (peak CAM-S =8-19), who experienced progressive decline to 3 years from a mean GCP of 53.8
234 at baseline to 51.8 at 36 months (2.0 average point decline).

235 Figure 1 shows the effect of GCP performance over time by delirium severity group. All
236 three groups experienced decline 1 month post-surgery and recovered to baseline or above. The
237 lowest severity group (peak CAM-S=0-2) had no significant decline over months 2-36 (-0.17
238 GCP units/year, 95% CI -0.35, 0.01). For the group with peak CAM-S=3-7, there was a
239 significant decrease in GCP score (-0.30 GCP units/year, 95% CI -0.51, -0.09). The magnitude of
240 this change was about a third of the change observed in the highest severity grouping, peak
241 CAM-S=8-19 (-0.82 GCP units/year, 95% CI -1.28, -0.37). These results suggest a graded
242 association of delirium severity and the rate of cognitive decline. Compared to patients in the
243 lowest severity group, the most severe delirium group demonstrated a 4.8-fold accelerated
244 decline (-0.82/-0.17). A linear trend test for differences in slope across severity group was
245 significant (p=0.009; Appendix). Moreover, the significant linear relationship between delirium
246 severity and GCP slope remained when peak CAM-S was considered as a continuous measure
247 (see Appendix).

248 Table 3 shows the prevalence of proxy-rated IQCODE impairment by delirium severity
249 group over time. Sample sizes differ between Table 3 and Table 2 because we could not always
250 locate or interview a suitable proxy informant for every surgical patient. For those in the low
251 severity group (peak CAM-S=0-2), there was no significant change in IQCODE impairment over
252 time. For the other severity groups, the prevalence of IQCODE impairment increased
253 significantly over time, with greater prevalence of IQCODE impairment with increasing delirium
254 severity (odds ratio [OR] 1.2 (95% CI 0.99, 1.5). Similar to the results for GCP, the association
255 with IQCODE impairment suggests a dose response (Figure 2 shows adjusted models), with the

256 strongest effect in the most severe group; however, the linear trend did not achieve statistical
257 significance (p=0.07).

258 In total, 103 participants experienced either death or nursing home placement between 6-
259 36 months. At baseline, these participants were older, fewer were married, had higher Charlson
260 comorbidity scores, more depressive symptoms, more ADL and IADL impairment, lower GCP
261 scores (see Appendix for detailed study sample description). They also had higher peak CAM-S
262 scores during hospitalization relative to the 457 participants who did not die and were not placed
263 in a nursing home. We observed increasing incidence across severity groups (15%, 20%, 28%
264 for peak CAM-S 0-2, 3-7, and 8-19, respectively) and a trend which approached but did not
265 achieve statistical significance (p=0.06) (see Appendix for additional details).

266 267 **DISCUSSION**

268 In this large prospective cohort of older persons without baseline dementia undergoing
269 elective surgery, patients experiencing higher delirium severity had greater rates of long-term
270 cognitive decline by serial neuropsychological testing (GCP). This finding was supported by
271 analyses examining the proxy IQCODE and risk of death or nursing home placement. These
272 findings suggest a dose-response effect where the risk of poor long-term outcomes increases
273 progressively across severity groups. The risk for greater cognitive decline was substantial and
274 statistically significant in the highest delirium severity grouping.

275 The findings utilizing the composite GCP measure demonstrated a 4.8-fold more rapid
276 decline between the highest and lowest severity groups. The per-year change in GCP in the long-
277 term (months 2-36) is about -0.17 GCP units/year, or -0.02 (-0.17/7.30) standard deviation (SD)
278 units/year in the lowest delirium severity group (peak CAM-S 0-2). Prior studies report declines
279 with cognitive aging in the absence of dementia to range between -0.01 and -0.04 SD units/year
280 [40-42]. Thus, patients with low delirium severity had a rate of cognitive decline (-0.02 SD per

281 year) comparable to previous studies for cognitively normal persons. By comparison, SAGES
282 patients with moderate severity declined by -0.30 GCP units/year (-0.04 SD units) and those with
283 the most severe delirium declined by -0.82 GCP units/year (-0.11 SD units). Our findings align
284 with prior work in patients undergoing coronary-artery bypass grafting in which the pattern of
285 cognitive decline is predicted by early postoperative cognitive decline (POCD) [42],
286 underscoring similarities in the long-term trajectories of patients with POCD and severe
287 postoperative delirium.

288 While the substantial short-term adverse outcomes of delirium are well-recognized, our
289 results hold important implications for the longer-term prognosis of delirium. This represents a
290 paradigm shift in the way delirium is currently viewed. Delirium may not be transient and
291 reversible with only acute complications; rather, more severe delirium cases may be
292 associated with long-term and potentially permanent cognitive decline. Furthermore, this work
293 suggests the need to target patients with high delirium severity for strategies to prevent
294 progressive cognitive decline, and potentially increased risk for dementia.

295 While prior work has established the association of incident delirium with long-term
296 cognitive decline [7-10], these findings are novel in demonstrating that delirium severity is
297 directly associated with long-term cognitive decline in an exposure-response fashion. We
298 acknowledge that causal associations cannot be determined from this observational study.
299 However, the observed exposure-response relationship is a critical first step in demonstrating a
300 direct association between delirium severity and long-term cognitive decline, and is an important
301 criterion used in causal inference for epidemiologic studies [43]. The novelty of our study also
302 includes both the use of a comprehensive measure of delirium severity (peak CAM-S scores,
303 reflecting the height of delirium intensity) and in the serial measurement of cognitive function
304 over a 3-year period following surgery. We chose peak CAM-S as our outcome measure to

305 reflect maximal intensity of delirium; however, other measures might have been chosen (e.g.,
306 sum CAM-S [18], see Appendix). Future studies should examine other severity measures,
307 including the Memorial Delirium Assessment Scale, Delirium Rating Scale, and Delirium Index
308 have been associated with increased mortality [44, 45], institutionalization [46, 47], and length
309 of stay [48]. Delirium duration has also been associated with increased death rates, increased
310 ventilator-dependent days, and intensive care unit stay [49-52]. The current study is innovative
311 in enabling examination of exposure-response relationships by examining outcomes across
312 multiple levels of severity. Other strengths include the use of a large cohort with thorough data
313 collection, careful characterization of preoperative cognition, repeated neuropsychological
314 testing over time, standardized delirium assessments, and extended post-surgical follow-up.
315 Additionally, exclusion of mild dementia at baseline facilitated examination of the effects of
316 delirium severity free of this potentially confounding influence. This presented a unique
317 opportunity to study cognitive impairment following delirium occurring largely in non-
318 cognitively impaired older patients. Finally, the careful correction for learning effects over time
319 represents another important advance.

320 Several caveats about this study deserve mention. Although we controlled for learning
321 effects, patients recovered back to or above baseline levels at 2 months, suggesting that: 1)
322 longer-term follow-up is critical to understanding the trajectory of cognitive recovery post-
323 surgery, and 2) this control for learning effects was either incomplete or that patients had
324 depressed cognitive levels at baseline, which may have been due to preadmission pain
325 medications such as narcotics. We encountered missing data due to deaths and drop-outs, and
326 addressed these in sensitivity analyses to assure the robustness of our conclusions (Appendix).
327 Despite using reasonable and established methods, participants who developed delirium may
328 have been on a downward cognitive trajectory prior to surgery, and we could not completely rule

329 out preclinical (asymptomatic) dementia, or clinically presymptomatic, but AD biomarker
330 positive dementia (as defined by stage 1 of the 2011 NIA criteria for AD), at baseline. Moreover,
331 the observation of a lower GCP in this group was anticipated, given that baseline cognitive
332 impairment has been long recognized as an important risk factor for delirium. Perhaps the more
333 intriguing observation is that participants on average improved back to baseline at 2 months
334 following delirium, and successively declined from 2 to 36 months suggesting a degree of initial
335 resiliency that would not be expected for those with underlying dementia. Similarly, we
336 acknowledge that inclusion of the pending follow-up visits may influence our current findings. In
337 general, we do not anticipate a substantial change in our study conclusions upon incorporating
338 the remaining visits since GCP scores observed for the two lowest delirium severity groups (peak
339 CAM-S 0-2 and 3-7) are relatively stable from around month 24 and onwards, and the GCP
340 scores appear to continue declining in the highest delirium severity group (peak CAM-S 8-19).
341 An additional caveat includes the fact that patients with delirium had lower GCP scores at
342 baseline than those without delirium, although both groups were above the U.S. population mean
343 GCP score=50. It may be that patients who were undergoing cognitive decline prior to surgery
344 may represent individuals at greatest risk for experiencing more severe delirium; however, with
345 only one preoperative cognitive assessment, we were unable to directly test this possibility. We
346 attempted to investigate this possibility by matching patients in the highest severity group (peak
347 CAM-S 8-19) with patients in the other two severity groups on preoperative GCP (see Appendix
348 for Methods and detailed Results), and found the pace of decline was faster in the highest
349 severity group (peak CAM-S 8-19; slope -0.09 SD/year) than in the peak CAM-S 3-7 group
350 (slope -0.04 SD/year), which was in turn faster than the peak CAM-S 0-2 group (slope -0.02
351 SD/year). We acknowledge that the study population represents a highly educated sample with
352 relatively low racial diversity from a single city; however, the diversity characteristics of our

353 sample (92% white) are representative of the Boston area (2008-2012 census data) [53]. It is
354 important to note that our choice of a dementia-free, relatively robust elective surgical population
355 may have influenced our findings. Patients with dementia might be more vulnerable to decline
356 after milder cases of delirium [5]. Finally, our use of the peak CAM-S does not discern
357 hypoactive from hyperactive delirium, which may have differing prognoses.

358 While delirium has previously been considered a transient condition of only short-term
359 significance, our results suggest that for patients with moderate to severe delirium, the declines
360 in cognition may be both substantial and long-term, and most notably exceeds the rate of decline
361 observed for patients with dementia. Although it remains critical to prevent and treat all delirium
362 to minimize well-documented short-term adverse outcomes, our results suggest the need for
363 more targeted strategies (e.g., cognitive rehabilitation, as used for patients with brain injuries
364 [54]) in patients with higher delirium severity to prevent long-term cognitive decline. Our
365 findings underscore the need to heighten efforts to better understand the risk factors and
366 pathophysiology of delirium of moderate to high severity, and to better target prevention and
367 management strategies to mitigate the long-term and potentially permanent adverse sequelae
368 associated with this common, morbid, and costly geriatric syndrome.

369 **Acknowledgement**

370

371 The authors thank the patients, family members, nurses, physicians, and research staff who
372 participated in the SAGES Study. This work is dedicated to the memory of Joshua Bryan Inouye
373 Helfand.

374

375

376

377

378

379

380

381 **Conflict of Interest Disclosures**

382 The authors state no conflicts of interest to report.

383

384

385

References

- 386
387
388 [1] Cole MG, Primeau FJ (1993) Prognosis of delirium in elderly hospital patients. *CMAJ*
389 **149**, 41-46.
- 390 [2] Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P (1998) Does delirium
391 contribute to poor hospital outcomes? A three-site epidemiologic study. *J Gen Intern*
392 *Med* **13**, 234-242.
- 393 [3] Inouye SK, Westendorp RG, Saczynski JS (2014) Delirium in elderly people. *Lancet* **383**,
394 911-922.
- 395 [4] Saczynski JS, Marcantonio ER, Quach L, Fong TG, Gross A, Inouye SK, Jones RN
396 (2012) Cognitive trajectories after postoperative delirium. *N Engl J Med* **367**, 30-39.
- 397 [5] Fong TG, Jones RN, Shi P, Marcantonio ER, Yap L, Rudolph JL, Yang FM, Kiely DK,
398 Inouye SK (2009) Delirium accelerates cognitive decline in Alzheimer disease.
399 *Neurology* **72**, 1570-1575.
- 400 [6] Gross AL, Jones RN, Habtemariam DA, Fong TG, Tommet D, Quach L, Schmitt E, Yap
401 L, Inouye SK (2012) Delirium and long-term cognitive trajectory among persons with
402 dementia. *Arch Int Med* **172**, 1-8.
- 403 [7] Inouye SK ME, Kosar CM, Tommet D, Schmitt EM, Trivison TG, Saczynski JS, Ngo
404 LH, Alsop DC, Jones RN (2016) The short- and long-term relationship between delirium
405 and cognitive trajectory in older surgical patients. *Alzheimers Dement* **12**, 766-775.
- 406 [8] Davis DH, Muniz Terrera G, Keage H, Rahkonen T, Oinas M, Matthews FE,
407 Cunningham C, Polvikoski T, Sulkava R, MacLulich AM, Brayne C (2012) Delirium is
408 a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain*
409 **135**, 2809-2816.

- 410 [9] Pandharipande PP, Girard TD, Jackson JC, Morandi A, Thompson JL, Pun BT, Brummel
411 NE, Hughes CG, Vasilevskis EE, Shintani AK, Moons KG, Geevarghese SK, Canonico
412 A, Hopkins RO, Bernard GR, Dittus RS, Ely EW, BRAIN-ICU Study Investigators
413 (2013) Long-term cognitive impairment after critical illness. *N Engl J Med* **369**, 1306-
414 1316.
- 415 [10] Girard TD, Yende S (2016) Cognitive impairment and critical illness: A chicken and an
416 egg. *Crit Care Med* **44**, 2115-2116.
- 417 [11] Inouye SK, Kosar CM, Tommet D, Schmitt EM, Puelle MR, Saczynski JS, Marcantonio
418 ER, Jones RN (2014) The CAM-S: development and validation of a new scoring system
419 for delirium severity in 2 cohorts. *Ann Intern Med* **160**, 526-33.
- 420 [12] Schmitt EM, Marcantonio ER, Alsup DC, Jones RN, Rogers SO Jr, Fong TG, Metzger E,
421 Inouye SK, SAGES Study Group (2012) Novel risk markers and long-term outcomes of
422 delirium: The Successful Aging after Elective Surgery (SAGES) Study design and
423 methods. *J Am Med Dir Assoc* **13**, 818.
- 424 [13] Schmitt EM, Saczynski JS, Kosar CM, Jones RN, Alsup DC, Fong TG, Metzger E,
425 Cooper Z, Marcantonio ER, Trivison T, Inouye SK, Successful Aging after Elective
426 Surgery Study Group (2015) The Successful Aging After Elective Surgery Study: Cohort
427 Description and Data Quality Procedures. *J Am Geriatr Soc* **63**, 2463-2471.
- 428 [14] Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI (1990) Clarifying
429 confusion: the confusion assessment method. A new method for detection of delirium.
430 *Ann Intern Med* **113**, 941-948.
- 431 [15] Wei LA, Fearing MA, Sternberg EJ, Inouye SK (2008) The Confusion Assessment
432 Method: a systematic review of current usage. *J Am Geriatr Soc* **56**, 823-830.

- 433 [16] Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST, Jr., Leslie DL, Agostini JV (2005)
434 A chart-based method for identification of delirium: validation compared with
435 interviewer ratings using the confusion assessment method. *J Am Geriatr Soc* 53, 312-
436 318.
- 437 [17] Saczynski JS, Kosar CM, Xu G, Puelle MR, Schmitt E, Jones RN, Marcantonio ER,
438 Wong B, Isaza I, Inouye SK (2014) A tale of two methods: chart and interview methods
439 for identifying delirium. *J Am Geriatr Soc* 62, 518-524.
- 440 [18] Vasunilashorn SM, Marcantonio ER, Gou Y, Pisani MA, Trivison TG, Schmitt EM,
441 Jones RN, Inouye SK (2016) Quantifying the severity of a delirium episode throughout
442 hospitalization: The combined importance of intensity and duration. *J Gen Intern Med*
443 31, 1164-1171.
- 444 [19] Greenland S (1995) Dose-response and trend analysis in epidemiology: alternatives to
445 categorical analysis. *Epidemiology* 6, 356-365.
- 446 [20] Trenerry MR (1990) *Visual Search and Attention Test: VSAT*. Psychological Assessment
447 Resources.
- 448 [21] Brandt J (1991) The Hopkins Verbal Learning Test: Development of a new memory test
449 with six equivalent forms. *Clin Neuropsychologist* 5, 125-142.
- 450 [22] Wechsler D (1981) *WAIS-R manual: Wechsler adult intelligence scale-revised*.
451 Psychological Corporation.
- 452 [23] Spreen OB (1977) *Neurosensory Center Comprehensive Examination for Aphasia:*
453 *Manual of instructions*. NCCEA.
- 454 [24] Mack WJ, Freed DM, Williams BW, Henderson VW (1992) Boston Naming Test:
455 shortened versions for use in Alzheimer's disease. *J Gerontol* 47, 154-158.

- 456 [25] Teng EL, Chui HC (1987) The Modified Mini-Mental State (3MS) examination. *J Clin*
457 *Psychiatry* 48, 314-318.
- 458 [26] Jones RN, Rudolph JL, Inouye SK, Yang FM, Fong TG, Milberg WP, Tommet D,
459 Metzger E, Cupples LA, Marcantonio ER (2010) Development of a unidimensional
460 composite measure of neuropsychological functioning in older cardiac surgery patients
461 with good measurement precision. *J Clin Exp Neuropsychol* 32, 1041-1049.
- 462 [27] Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, Burke
463 JR, Fisher GG, Fultz NH, Hurd MD, Potter GG, Rodgers WL, Steffens DC, Weir DR,
464 Willis RJ (2005) The Aging, Demographics, and Memory Study: study design and
465 methods. *Neuroepidemiology* 25, 181-191.
- 466 [28] Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK (2014) Calibration and validation
467 of an innovative approach for estimating general cognitive performance.
468 *Neuroepidemiology* 42, 144-153.
- 469 [29] Cavallari M, Hshieh TT, Guttmann CR, Ngo LH, Meier DS, Schmitt EM, Marcantonio
470 ER, Jones RN, Kosar CM, Fong TG, Press D, Inouye SK, Alsup DC, SAGES Study
471 Group (2015) Brain atrophy and white-matter hyperintensities are not significantly
472 associated with incidence and severity of postoperative delirium in older persons without
473 dementia. *Neurobiol Aging* 36, 2122-2129.
- 474 [30] Saczynski JS, Inouye SK, Kosar CM, Tommet D, Marcantonio ER, Fong T, Hshieh T,
475 Vasunilashorn S, Metzger ED, Schmitt E, Alsup DC, Jones RN, SAGES Study Group
476 (2014) Cognitive and brain reserve and the risk of postoperative delirium in older
477 patients: analysis of data from a prospective observational study. *Lancet Psychiatry* 1,
478 437-443.

- 479 [31] Evered L, Scott DA, Silbert B, Maruff P (2011) Postoperative cognitive dysfunction is
480 independent of type of surgery and anesthetic. *Anesth Analg* **112**, 1179-1185.
- 481 [32] Lewis M, Maruff P, Silbert B (2004) Statistical and conceptual issues in defining post-
482 operative cognitive dysfunction. *Neurosci Biobehav Rev* **28**, 433-440.
- 483 [33] Soinne L, Helenius J, Tikkala I, Saimanene E, Salonen O, Hietanen M, Lindsberg PJ,
484 Kaste M, Tatlisumak T (2009) The effect of severe carotid occlusive disease and its
485 surgical treatment on cognitive functions of the brain. *Brain Cogn* **69**, 353-359.
- 486 [34] Jorm AF (1994) A short form of the Informant Questionnaire on Cognitive Decline in the
487 Elderly (IQCODE): development and cross-validation. *Psychol Med* **24**, 145-153.
- 488 [35] Yesavage J, Sheikh JI (1986) Geriatric Depression Scale (GDS) Recent Evidence and
489 Development of a Shorter Version. *Clin Gerontol* **5**, 165-173.
- 490 [36] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW (1963) Studies of Illness in the
491 Aged. The Index of ADL: A Standardized Measure of Biological and Psychosocial
492 Function. *JAMA* **185**, 914-991.
- 493 [37] Lawton MP, Brody EM (1969) Assessment of older people: self-maintaining and
494 instrumental activities of daily living. *Gerontologist* **9**, 179-186.
- 495 [38] Ware JE, Jr., Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36).
496 I. Conceptual framework and item selection. *Med Care* **30**, 473-483.
- 497 [39] Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying
498 prognostic comorbidity in longitudinal studies: development and validation. *J Chronic*
499 *Dis* **40**, 373-383.
- 500 [40] Hayden KM, Reed BR, Manly JJ, Tommet D, Pietrzak RH, Chelune GJ, Yang FM,
501 Revell AJ, Bennett DA, Jones RN (2011) Cognitive decline in the elderly: an analysis of
502 population heterogeneity. *Age Ageing* **40**, 684-689.

- 503 [41] Johnson JK, Gross AL, Pa J, McLaren DG, Park LQ, Manly JJ (2012) Longitudinal
504 change in neuropsychological performance using latent growth models: a study of mild
505 cognitive impairment. *Brain Imaging Behav* 6, 540-550.
- 506 [42] Newman MF, Kirchner JL, Phillips-Bute B, Gaver V, Grocott H, Jones RH, Mark DB,
507 Reves JG, Blumenthal JA. (2001) Longitudinal assessment of neurocognitive function
508 after coronary-artery bypass surgery. *N Engl J Med* 344, 395-402.
- 509 [43] Hill AB (1965) The environment and disease: Association or causation? *Proc Royal Soc*
510 *Med* 58, 295-300.
- 511 [44] Kelly KG, Zisselman M, Cutillo-Schmitter T, Reichard R, Payne D, Denman SJ (2001)
512 Severity and course of delirium in medically hospitalized nursing facility residents. *Am J*
513 *Geriatr Psychiatry* 9, 72-77.
- 514 [45] McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E (2002) Delirium predicts
515 12-month mortality. *Arch Intern Med* 162, 457-463.
- 516 [46] Marcantonio E, Ta T, Duthie E, Resnick NM (2002) Delirium severity and psychomotor
517 types: their relationship with outcomes after hip fracture repair. *J Am Geriatr Soc* 50,
518 850-857.
- 519 [47] Dasgupta M, Brymer C (2014) Prognosis of delirium in hospitalized elderly: worse than
520 we thought. *Int J Geriatr Psychiatry* 29, 497-505.
- 521 [48] Zhang W, Hu W, Shen M, Ye X, Huang Y, Sun Y (2016) Profiles of delirium and the
522 clinical outcomes of patients who underwent coronary artery bypass grafting: a
523 prospective study from China. *J Clin Nurs* 25, 631-641.
- 524 [49] Pisani MA, Kong SY, Kasl SV, Murphy TE, Araujo KL, Van Ness PH (2009) Days of
525 delirium are associated with 1-year mortality in an older intensive care unit population.
526 *Am J Respir Crit Care Med* 180, 1092-1097.

- 527 [50] Ely EW, Shintani A, Truman B, Speroff T, Gordon SM, Harrell FE Jr, Inouye SK,
528 Bernard GR, Dittus RS (2004) Delirium as a predictor of mortality in mechanically
529 ventilated patients in the intensive care unit. *JAMA* **291**, 1753-1762.
- 530 [51] Shehabi Y, Riker RR, Bokesch PM, Wisemandle W, Shintani A, Ely EW (2010)
531 Delirium duration and mortality in lightly sedated, mechanically ventilated intensive care
532 patients. *Crit Care Med* **38**, 2311-2318.
- 533 [52] Bellelli G, Mazzola P, Morandi A, Bruni A, Carnevali L, Corsi M, Zatti G, Zambon A,
534 Corrao G, Olofsson B, Gustafson Y, Annoni G (2014) Duration of postoperative delirium
535 is an independent predictor of 6-month mortality in older adults after hip fracture. *J Am*
536 *Geriatr Soc* **62**, 1335-1340.
- 537 [53] U.S. Census, Bureau. 2008-2012 American Community Survey.
538 <http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>, Accessed on August 30,
539 2016.
- 540 [54] Park HY, Maitra K, Martinez KM (2015) The effect of occupation-based cognitive
541 rehabilitation for traumatic brain injury: A meta-analysis of randomized controlled trials.
542 *Occup Ther Int* **22**, 104-116.
- 543 [55] National Research Council Panel on Handling Missing Data in Clinical Trials (2010) The
544 Prevention and Treatment of Missing Data in Clinical Trials. National Academies Press,
545 Washington DC.

546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569

SAGES Study Group

[Presented in alphabetical order; individuals listed may be part of multiple groups, but are listed only once under major activity, listed in parentheses].

Overall Principal Investigator: Sharon K. Inouye, MD, MPH (Overall PI, Administrative Core, Project 1; HSL, BIDMC, HMS).

Project and Core Leaders: David Alsop, PhD (Project 3; BIDMC, HMS); Richard Jones, ScD (Data Core, Project 4; Brown University); Thomas Trivison, PhD (Data Core, HSL, HMS); Edward R. Marcantonio, MD, SM (Overall Co-PI, Epidemiology Core, Project 2; BIDMC, HMS).

Executive Committee: Steven Arnold, MD, (MGH); Zara Cooper, MD, MSc (HMS, BWH); Bradford Dickerson, MD (MGH, HMS); Tamara Fong, MD, PhD (HMS, HSL, BIDMC); Towia Libermann, PhD (HMS, BIDMC); Eran Metzger, MD, (HMS, HSL, BIDMC); Alvaro Pascual-Leone, MD (HMS, BIDMC); Eva M. Schmitt, PhD (Overall Project Director, HSL); Mouhsin Shafi, MD (HMS, BIDMC).

Other Co-investigators: Michele Cavallari, MD (BWH); Weiyang Dai, PhD (BIDMC); Simon T. Dillon, PhD (HMS, BIDMC); Janet McElhaney, MD (UConn); Charles Guttman, MD (BWH, HMS); Tammy Hsieh, MD (BWH); George Kuchel, MD, FRCP, (UConn); Long Ngo, PhD (HMS, BIDMC); Daniel Press, MD (HMS, BIDMC); Jane Saczynski, PhD, (UMASS); Sarinnapha Vasunilashorn, PhD (HMS, BIDMC).

Clinical Consensus Panel: Margaret O'Connor, PhD (HMS, BIDMC); Eyal Kimchi, MD, PhD (MGH), Jason Strauss, MD (Cambridge Health Alliance); Bonnie Wong, PhD (BIDMC).

Surgical Leaders: Michael Belkin, MD (HMS, BWH); Douglas Ayres, MD (HMS, BIDMC); Mark Callery, MD (HMS, BIDMC); Frank Pomposelli, MD (HMS, BIDMC); John Wright, MD (HMS, BWH); Marc Schermerhorn, MD (HMS, BIDMC).

570 Epidemiology Core: Asha Albuquerque (HSL); Amanda Brown M.Ed. (HSL); Amy Callahan
571 (BIDMC), Sarah Dowal, MSW, LCSW, MPH (HSL); Meaghan Fox (BIDMC); Jacqueline
572 Gallagher, MS (BIDMC); Rebecca Anna Gersten; Ariel Hodara (BIDMC); Ben Helfand, MPH
573 (BIDMC); Jennifer Inloes (HSL); Jennifer Kettell (HSL); Aleksandra Kuczmarska (BIDMC);
574 Jacqueline Nee (HSL); Emese Nemeth (HSL); Lisa Ochsner (BWH); Kerry Palihnich (BIDMC);
575 Katelyn Parisi (HSL); Margaret Puelle (HSL); Sarah Rastegar, MA (HSL); Margaret Vella
576 (HSL), Guoquan Xu, MD, PhD (HSL).

577 Data Management and Statistical Analysis Core: Margaret Bryan (HSL); Jamey Guess
578 (BIDMC); Dee Enghorn (HSL); Alden Gross, PhD, MHS (John Hopkins School of Medicine);
579 Yun Gou, MA (HSL); Daniel Habtemariam (HSL); Ilean Isaza, PhD (HSL); Cyrus Kosar, MA
580 (HSL); Christopher Rockett, PhD (HSL); Douglas Tommet, MPH (Brown University).

581 Fiscal Management Committee: Ted Gruen (HSL); Meg Ross (HSL); Katherine Tasker
582 (Chair, HSL).

583 Scientific Advisory Board: James Gee, PhD (University of Pennsylvania); Ann Kolanowski,
584 PhD, RN, FAAN (Pennsylvania State University); Margaret Pisani, MD, MPH (Yale
585 University); Sophia de Rooij, MD, PhD (Academic Medical Center, Amsterdam); Selwyn
586 Rogers, MD, MPH (Temple University), Stephanie Studenski, MD (Chair, NIA); Yaakov Stern,
587 PhD (Columbia University); Anthony Whittmore, MD (BWH, HMS).

588 Internal Advisory Board: Gary Gottlieb, MD, MBA (BWH, MGH, HMS); John Orav, PhD
589 (BWH, HMS); Reisa Sperling, MD, MMSc (BWH, HMS).

590 Abbreviations: BIDMC, Beth Israel Deaconess Medical Center; BWH, Brigham and Women's Hospital; HMS,
591 Harvard Medical School; HSL, Hebrew SeniorLife; MGH, Massachusetts General Hospital; PI, principal
592 investigator; UCONN, University of Connecticut Health Center.

593
594

Table 1. Description of Study Sample

Characteristic	Peak CAM-S score				Rank Correlation ^c
	Full Sample (N = 560)	0 – 2 (N = 244)	3 – 7 (N = 248)	8-19 (N = 68)	
				77.0 (4.6)	0.09
Age - mean (SD)	76.7 (5.2)	76.0 (4.7)	77.2 (5.7)	(4.6)	
Female – n (%)	326 (58)	147 (60)	141 (57)	38 (56)	-0.04
Nonwhite – n (%)	42 (8)	12 (5)	25 (10)	5 (7)	0.07
				14.6	-0.17
Education – mean years (SD)	15.0 (2.9)	15.6 (2.8)	14.4 (2.9)	(3.0)	
Married – n (%)	332 (59)	142 (58)	151 (61)	39 (57)	0.01
Lives Alone – n (%)	167 (30)	79 (32)	66 (27)	22 (32)	-0.03
Charlson score - n (%)					0.12
0	257 (46)	126 (52)	102 (41)	29 (43)	
1	139 (25)	62 (25)	66 (27)	11 (16)	
2+	164 (29)	56 (23)	80 (32)	28 (41)	
GDS15 score - n (%)					0.18
0 - 5	489 (88)	225 (93)	214 (86)	50 (74)	
6 - 15	69 (12)	17 (7)	34 (14)	18 (26)	
				53.8	-0.36
GCP score - mean (SD)	57.6 (7.3)	60.5 (6.7)	55.8 (7.3)	(5.6)	
3MS score - n (%)					0.13
85-100	523 (93)	237 (97)	225 (91)	61 (90)	
71-84	37 (7)	7 (3)	23 (9)	7 (10)	
Proxy IQCODE (baseline) - n (%)					0.104
Not Impaired	430 (78)	198 (83)	183 (76)	49 (72)	
Impaired	118 (22)	40 (17)	59 (24)	19 (28)	
ADL impairment – n (%)	42 (8)	10 (4)	24 (10)	8 (12)	0.12
IADL impairment – n (%)	152 (27)	51 (21)	77 (31)	24 (35)	0.13
Surgery type - n (%)					-0.03
Orthopedic	454 (81)	196 (80)	201 (81)	57 (84)	
Vascular	35 (6)	11 (5)	18 (7)	6 (9)	
General	71 (13)	37 (15)	29 (12)	5 (7)	
Delirium ^b - n (%)					
None	426 (76)	243 (100)	181 (73)	2 (3)	
Delirium	134 (24)	1 ^a (0)	67 (27)	66 (97)	

^aThe patient, with a peak CAM-S score of 2, had chart delirium

^b Delirium status was determined with daily interviews rating the Confusion Assessment Method, augmented by a validated chart review

^cSpearman rank correlation coefficient indicates the correlation of each variable with the peak CAM-S score

ADL = Activities of Daily Living, impairment indicated by human assistance to complete any activity

CAM-S = Confusion Assessment Method-Severity

GCP = General Cognitive Performance, composite measure of neuropsychological measures reflecting cognitive domains vulnerable to delirium, see text for details

GDS15= Geriatric Depression Scale 15 point version, range (0-15), higher is worse; a score 6 and above is considered impaired

IADL = Instrumental Activities of Daily Living, impairment indicated by human assistance to complete any activity

IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly, score >3.2 indicates cognitive impairment

3MS = Modified Mini-Mental State Exam, range (0-100), lower score indicates impairment; a score ≤84 is considered impaired

SAGES = Successful Aging after Elective Surgery Study

SD= standard deviation.

The Charlson comorbidity score ranged from 0-35, with higher scores indicating more comorbidity.

Table 2: Corrected GCP Scores over Time

Visit month	Full Sample		Peak CAM-S score					
	N	Mean (SD)	0 – 2		3 – 7		8 – 19	
			N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
0	560	57.6 (7.3)	244	60.5 (6.7)	248	55.8 (7.3)	68	53.8 (5.6)
1	548	56.8 (7.9)	243	60.0 (6.9)	242	55.0 (7.9)	63	51.4 (5.8)
2	536	58.0 (7.9)	238	60.9 (7.1)	237	56.2 (8.1)	61	53.8 (5.3)
6	528	58.2 (7.5)	237	61.0 (6.5)	230	56.4 (7.9)	61	54.2 (6.1)
12	511	58.4 (7.6)	227	61.2 (7.0)	224	56.8 (7.6)	60	53.9 (5.4)
18	499	58.3 (8.0)	219	61.5 (6.9)	222	56.5 (7.9)	58	52.7 (7.2)
24	474	58.2 (8.0)	213	61.2 (6.8)	211	56.4 (8.1)	50	52.4 (7.2)
30	325	57.5 (8.2)	132	60.7 (7.5)	152	56.1 (7.9)	41	52.4 (7.3)
36	312	57.1 (8.4)	123	60.6 (7.4)	141	55.8 (8.2)	48	51.8 (7.7)

CAM-S = Confusion Assessment Method-Severity

GCP = General Cognitive Performance, composite measure of neuropsychological measures reflecting cognitive domains vulnerable to delirium, see text for details

Notes: All postoperative GCP values corrected for practice effects (see text for details). The number of participants completing each the interview/the number of participants eligible for the interview for each time point follows with amount of attrition from the prior time point in brackets. Baseline: 560/560 [0]; Month 1: 548/552 [8]; Month 2: 536/546 [6]; Month 6: 528/539 [7]; Month 12: 511/527 [8]; Month 18: 499/516 [8]; Month 24: 474/489 [13]; Month 30: 325/342 [6]; Month 36: 312/316 [1]

Table 3: Empirically Observed Prevalence of Proxy IQCODE Impairment over Time

Visit month	Full Sample		Peak CAM-S score					
	N	n (%)	0 – 2		3 – 7		8 – 19	
			N	n (%)	N	n (%)	N	n (%)
0	548	118 (22)	238	40 (17)	242	59 (24)	68	19 (28)
6	514	135 (26)	229	46 (20)	226	67 (30)	59	22 (37)
12	487	130 (27)	217	49 (23)	218	60 (28)	52	21 (40)
18	480	142 (30)	208	49 (24)	217	66 (30)	55	27 (49)
24	452	125 (28)	202	46 (23)	205	61 (30)	45	18 (40)
30	314	101 (32)	127	28 (22)	145	54 (37)	42	19 (45)
36	287	94 (33)	118	25 (21)	126	48 (38)	43	21 (49)

CAM-S = Confusion Assessment Method-Severity

IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly

Since the IQCODE is proxy-rated, the sample sizes in this table reflect the availability of proxy-informants over time; 12 patients did not have any proxies available at baseline, yielding a total proxy sample of N=548

N=total possible sample, n=number with proxy IQCODE impairment

Figure 1

Title: Trajectory of General Cognitive Performance by Estimated Peak Confusion Assessment Method-Severity (CAM-S) Score

Legend: Figure 1 demonstrates the relationship between estimated general cognitive performance (GCP) and time following surgery (months) by delirium severity group. The model is adjusted for baseline GCP, age, gender, non-white race, education, Charlson score, Geriatric Depression Scale score, instrumental Activities of Daily Living (IADL) impairment, surgery type, and proxy Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) impairment. For each group, we plot the model-implied trajectory and a solid gray reference line at the baseline value. The amount of punctuation (acute decline at one month), recovery (up to two months), and long-term decline (two to 36 months) is shown by each CAM-S severity group, 0-2 (dashed black line), 3-7 (dot-dashed black line) and 8-19 (solid gray line). In the acute (punctuation) phase, all groups decline with the most severe group declining the most. This is followed by recovery of cognitive performance, with the less severe groups recovering (at two months) past their baseline (0 months) GCP score, and those in the most severe group showing an incomplete return to baseline. Over long-term follow-up, the less severe groups gradually decline in GCP performance, whereas the most severe group demonstrates a faster pace of decline.

Figure 1

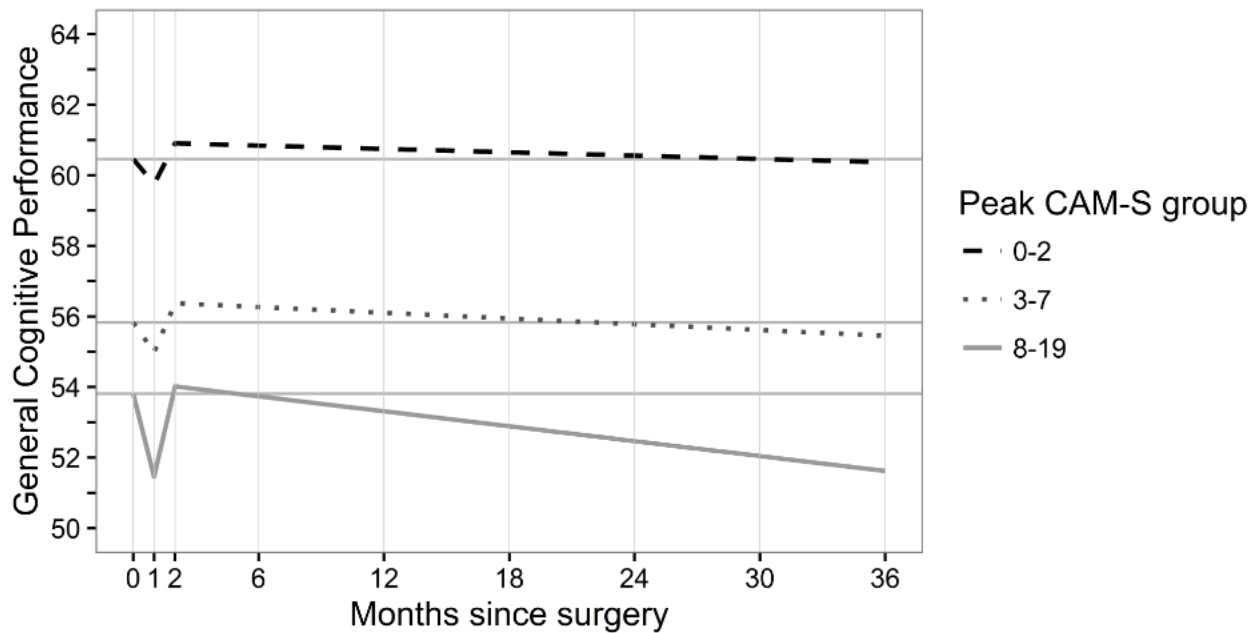
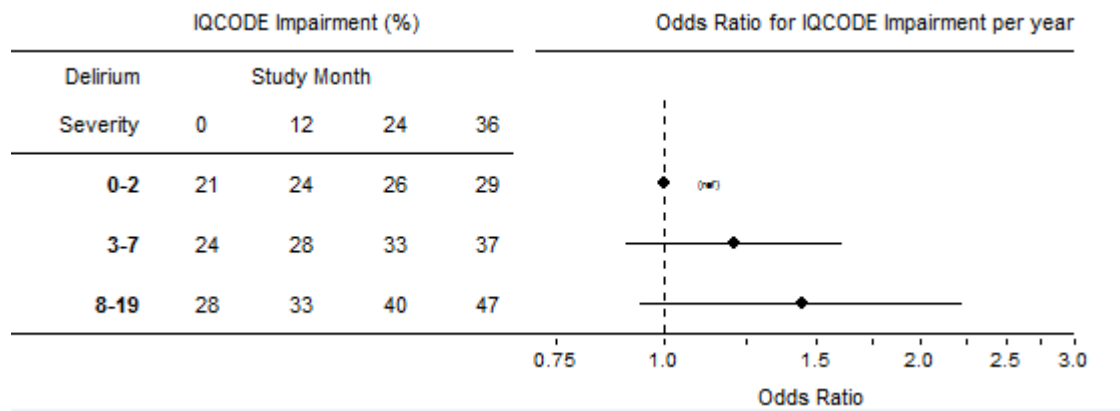


Figure 2

Title: Predicted Prevalence of IQCODE impairment by delirium severity group and study month

Legend: Figure 2 demonstrates the relationship between the prevalence of proxy Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) impairment (score ≥ 3.2) over time (study month) following surgery, calculated using a mixed effects generalized linear model. The odds ratios (OR) are computed from models that controlled for age, gender, non-white race, education, Charlson score, Geriatric Depression Scale score, and surgery type; and thus differ from the ORs derived from the numbers presented in Table 3. Model-implied (or expected) proportions with IQCODE impairment given mean values on covariates are presented in the table. Odds ratios (and 95% confidence bands) illustrate the size and precision of estimates of the delirium severity group by time (in years following surgery) interaction effects. Over time, all groups have increasing probability of being classified as impaired on the IQCODE ($p=.05$). The per-year odds of IQCODE ≥ 3.2 for this group is about two times greater than that observed for the lowest delirium severity group.



Vasunilashorn et al., Delirium Severity Post-Surgery and its Relationship with Long-term Cognitive Decline in a Cohort of Patients without Dementia

Appendix

Table of Contents	
	Page
Overview	32-33
Consideration of Practice and Retest Effects	34
Statistical Modeling Approach for Manuscript Analysis: General Cognitive Performance (GCP) as an Outcome	
<u>Figure S1.</u> A Path Diagram Illustrating the Estimated Model	35-36
<u>Table S1.</u> Detailed Model Results Reporting the Association between Delirium Severity (Peak Confusion Assessment Method-Severity [CAM-S]) and GCP	37
<i>Sensitivity Analysis 1: Assessing Robustness of Findings to Extreme Assumptions of Missing Data Due to Drop-Out, Death, or Institutionalization</i>	
<u>Table S2.</u> Peak CAM-S scores by study status	38-39
<u>Table S3.</u> Detailed Model Results Reporting the Association between Delirium Severity (peak CAM-S) and GCP, Given Extreme Assumptions of Missing Data	40
Interpretation of Results for Sensitivity Analysis 1	41
<i>Sensitivity Analysis 2: Assessing the Relationship Between Sum of CAM-S Scores and Long-Term Cognitive Decline</i>	
<u>Table S4.</u> Relationship of Peak CAM-S to Sum of All CAM-S Scores	42
<u>Table S5.</u> Sum of CAM-S Scores by Corrected GCP Scores Over Time	43
<u>Table S6.</u> Empirically Observed Prevalence of Proxy IQCODE Impairment Over Time	44
<u>Figure S2.</u> Trajectories of Estimated GCP by Sum of CAM-S Groups	45
<u>Figure S3.</u> Predicted Prevalence of Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) Impairment by Sum of CAM-S Group and Study Month	46
<u>Table S7.</u> Detailed Model Results Reporting the association between Delirium Severity (Sum CAM-S) and GCP	47
Analytic sample and follow-up rates	
<u>Figure S4.</u> STROBE Diagram of the Analytic Sample	48
<u>Table S8.</u> SAGES Follow-up Success Rate	49
Death or Nursing Home Placement Analysis	
<u>Table S9.</u> Description of Study Sample by Death or Nursing Home Placement Status	50-51
CAM-S Distribution	
<u>Figure S5.</u> Continuous Peak CAM-S Measure by GCP slope	52
Matched analysis	
<u>Table S10.</u> Results of Matched Analysis	53
References	54

Overview

The information in this Supplemental Appendix is provided to: (1) describe our approach for minimizing learning and practice (retest) effects; (2) describe our statistical modeling approach to test our hypotheses regarding the association of delirium and cognitive change following surgery with detailed model results; (3) assess the limits of our inferences in the presence of differing assumptions about missing data [sensitivity analysis 1]; (4) assess the relationship between long-term cognitive decline and an alternate measure of delirium severity, sum of CAM-S scores [sensitivity analysis 2]; (5) describe our analytic sample and follow-up success rates; (6) describe the study sample by death or nursing home placement status; (6) illustrate the association between peak CAM-S as a continuous variable and GCP slope; and (7) consideration of a matched analysis

Consideration of Practice and Retest Effects

Practice and retest effects are pervasive and challenging in studies with repeated measures of cognition over time. No clear consensus exists for the optimal handling of practice (and retest) effects in statistical analyses.¹ The technique we uses was first used in the International Study on Postoperative Cognitive Dysfunction (ISPOCD), and as implemented here is quite similar to the modeling of boost retest effects, described previously,² which lends further support for its application here. Our approach involves assessing a comparison sample (n = 119) of otherwise comparable persons (patients in a primary care clinic at one of our study sites) with the same tests and on the same schedule of assessment. The mean performance of this sample at each time point is used to center the observed scores seen in our surgical sample at matching time points. The 6 month assessment in our comparison sample is used as the centering point for all subsequent observations in our surgical sample. The 6 month cutoff was utilized, since most studies consider that practice effects have leveled off by this time. This approach relies upon the assumption that differences in the mean across the repeat performances in the comparison sample represent the mean practice or retest effect free of normative cognitive change. We considered this assumption reasonable given the very short time interval between assessments in a relatively healthy comparison group.

Statistical Model

Our statistical modeling approach involved the use of generalized linear mixed effects models, or random effects models, that account for: (1) the dependence of pre-operative cognitive performance (y_0) on preoperative baseline and background variables (z), (2) the dependence of the severity level of delirium (d_c) on background variables (z) and pre-operative level of cognitive functioning (y_0), and (3) the dependence of follow-up cognitive performance (y_t) on baseline (preoperative) cognitive performance (y_0), delirium (d_c) and background variables (z). A graph summarizing the temporal ordering (left to right) and dependence relationships is shown in Figure S1.

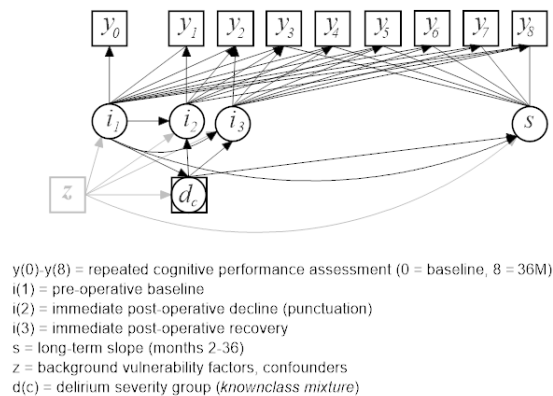


Figure S1. A path diagram illustrating the estimated model. Variables include background and pre-operative baseline variables considered as potential confounders (z), baseline pre-operative cognitive performance (General Cognitive Performance, GCP; y_0), the occurrence and level of delirium severity following surgery (d_c), and follow-up cognitive performance (through the scheduled 36 month follow-up; y_1 - y_8). Change in GCP (y) is modeled with a piecewise linear latent growth curve model, where the “pieces” refer to “boost” effects at one and two months capturing initial decline and recovery following surgery (which occurs between y_0 and y_1). Latent variables (enclosed in circles) capture baseline (i_1), and initial decline, recovery, and long-term slope (i_2 , i_3 , s), as well as a categorical known class latent mixture variable (d_c) that is identical to observed d_c , the post-operative delirium severity group. We regress d_c on i with multinomial logistic regression, which captures the dependency of delirium and its severity on baseline cognitive function. We also regress baseline cognition and delirium severity group, as well as the three slope pieces on background and potentially confounding variables. Finally, we also regress the immediate decline, recovery, and long-term slope effects on baseline.

We are primarily interested in the direct effect of delirium severity group (d_c) on follow-up GCP, General Cognitive Performance (y_t). The estimated model is one that includes linear regression, multinomial logistic regression, and piecewise linear mixed effect regression models, all estimated simultaneously in a multiple group or known class mixture model. Baseline cognitive performance regressed on background variables are handled with linear regression. The regression of postoperative delirium severity group on preoperative cognition and background variables is handled with logistic regression, as the outcome membership in severity group (d_c) is a set of nominal outcomes (0 = in lowest severity class, 1 = in severity class c). The piecewise linear mixed effect model for follow-up cognition includes two (fixed) pieces for performance at 1 and 2 months, parameter estimates for which describe the punctuation and recovery effects, respectively, following surgery and/or delirium. Change over time from scheduled study month 2 through 36 was modeled with a mixed effect model, with change over time included as a linear random effect. We considered a quadratic effect to model follow-up time but observed only a small improvement to information criteria (difference in Bayesian information criterion <2 , considered an insignificant effect⁴), and the difference in the Akaike information was criteria $<1\%$. Thus, we considered the gain in explanatory power insufficient to justify the added complexity.

Parameter estimates were obtained with Mplus software version 7.4 (Muthén & Muthén, Los Angeles, CA) using robust maximum likelihood (ML) parameter estimation. The mixed effect model included random effects for preoperative baseline and linear slope, meaning that

these parameters were modeled with variances and covariances. Piecewise effects for punctuation and retest were modeled as fixed effects. The modeling approach does not make explicit reference to the covariance structure of the repeatedly observed outcome (cognitive performance, y_0, y_t) as in other mixed effect modeling approaches, but can be conceptualized as *unstructured* and accounted for by the variances in baseline, linear slope, and also by background variables and delirium severity group (z, d_c). ML parameter estimation makes use of all available information, and the parameter estimates reflect the most likely parameter estimates for persons with incomplete data over the follow-up period. We had missing data for 2 participants for one background variable (Geriatric Depression Scale, GDS score). The other background variables (z , age, gender, non-white race, education, Charlson score, instrumental activities of daily living [IADL], and surgery type) or postoperative delirium (d_c) did not have missing data. Bayesian estimation methods were used to impute values for the missing data.

Detailed Model Results

Below, we provide detailed model results, expanded beyond what is displayed in the main manuscript. We omit the effects of covariates here, since all covariates were mean-centered and therefore do not influence the computation of expected values. General cognitive performance (GCP) is centered at the overall sample mean in the model, and estimates in all tables are transformed to reflect the full scale of the GCP.

Table S1. Detailed Model Results Reporting the Association between Delirium Severity (Peak Confusion Assessment Method-Severity score) and General Cognitive Performance (N=560)

Model Parameter	Estimate	95% CI	P-value
Effect of baseline GCP in ...			
GCP punctuation	0.53	(0.11, 0.93)	.013
GCP recovery	-0.11	(-0.47, 0.26)	.565
GCP slope, Months 2 – 36 (per year)	0.18	(0.01, 0.37)	.063
Estimated means for latent growth model effects by Delirium (peak CAM-S) severity group			
CAM-S Peak 0-2			
GCP Punctuation	-0.65	(-1.06, -0.23)	.002
GCP Recovery	1.15	(0.79, 1.50)	<.001
GCP Slope, Months 2 - 36 (per year)	-0.17	(-0.35, 0.01)	.073
CAM-S Peak 3-7			
GCP Punctuation	-0.78	(-1.19, -0.28)	<.001
GCP Recovery	1.37	(0.99, 1.76)	<.001
GCP Slope, Months 2 - 36 (per year)	-0.30	(-0.51, -0.09)	.005
CAM-S Peak 8-19			
GCP Punctuation	-2.28	(-3.25, -1.31)	<.001
GCP Recovery	2.55	(1.78, 3.31)	<.001
GCP Slope, Months 2 - 36 (per year)	-0.82	(-1.28, -0.37)	<.001

SAGES = Successful Aging after Elective Surgery, GCP = general cognitive performance, CAM-S = Confusion Assessment Method-Severity score, CI = confidence interval

Sensitivity Analysis 1: Assessing Robustness of Findings to Extreme Assumptions of Missing Data Due to Drop-Out, Death, or Institutionalization

There was incomplete follow-up data through the 36 month follow-up visit. Most of this is due to the rolling enrollment of the study design. However, there were some cases of death and dropout.

Table S2. Peak Confusion Assessment Method-Severity scores by study status

Study status	CAM-S Peak Scores							
	Overall		0-2		3-7		8-19	
	n	%	n	%	n	%	n	%
In study	496	(89)	225	(92)	218	(88)	53	(78)
Death/dropout	64	(11)	19	(8)	30	(12)	15	(22)
Total	560		244		248		68	

CAM-S = Confusion Assessment Method-Severity

The difference in proportion across all delirium severity groups is significant ($p = .004$). Our main analysis reported in the manuscript summarizes maximum likelihood parameter estimates, which are theoretically unbiased under the assumption that the missing data mechanism is missing at random (MAR). This means that the reason why individuals are missing is not due to the value on the outcome (GCP) that would have been observed, had it been observed (conditional on the effect of observed data). Most of the missing data can be safely assumed to be missing completely at random (MCAR) because it is due to the date of enrollment and no other factor. However, it is possible that for some of the people who dropped out due to death or institutionalization the MAR assumption is overly restrictive. To address this, we performed a set of sensitivity analyses to examine the range of possible effects of a non-ignorable missing data pattern for those participants for whom we are assuming MAR holds. In both sensitivity analyses, we impute values for missing data. We do so under two conditions that represent extreme conditions of possible missing data mechanisms that would be most beneficial and most harmful to our hypothesis that delirium severity influences long-term cognitive decline. Both analyses follow a similar framework. Factor scores for each participant’s baseline, decline, recovery, and long-term slope were estimated from the adjusted model shown in the main manuscript. The long-term slope estimates were modified by either adding or subtracting an amount proportional to the standard deviation of the long-term slopes. The original baseline, decline, recovery, and the modified long-term slope factor scores were then used to calculate the missing outcomes for subjects that died or dropped out.

Best case scenario: This scenario provides conditions that are most favorable to our hypotheses. The long-term slope factor scores were modified so that the delirium severity groups would diverge. The participants in the peak Confusion Assessment Method-Severity (CAM-S) 0-2 group had 1 standard deviation (SD) added to their score. The participants in the CAM-S 3-7 group had 0.5 SD added to their score. The participants in the CAM-S 8-19 group had 0.5 SD subtracted from their score.

Worst case scenario: This scenario provides conditions that are least favorable to our hypotheses. The long-term slope factor scores were modified so that the delirium severity

groups would converge. The participants in the CAM-S 0-2 group had 1 SD subtracted from their score. The participants in the CAM-S 3-7 group had 0.5 SD subtracted from their score. The participants in the CAM-S 8-19 group had 0.5 SD added their score.

Table S3. Detailed Model Results Reporting the Association between Delirium Severity (peak Confusion Assessment Method-Severity score) and General Cognitive Performance, Given Extreme Assumptions of Missing Data

	Baseline		Decline		Recovery		Long-term slope	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Original model (for reference)								
Baseline GCP			0.53	(0.11, 0.93)	-0.11	(-0.47, 0.26)	0.18	(-0.01, 0.37)
CAM-S Peak:								
0-2	57.61	(57.12, 58.09)	-0.65	(-1.06, -0.23)	1.15	(0.79, 1.50)	-0.17	(-0.35, 0.01)
3-7	57.61	(57.12, 58.09)	-0.78	(-1.20, -0.35)	1.37	(0.99, 1.76)	-0.30	(-0.51, -0.09)
8-19	57.61	(57.12, 58.09)	-2.28	(-3.25, -1.31)	2.55	(1.78, 3.31)	-0.82	(-1.28, -0.37)
Sensitivity analysis: best case scenario								
Baseline GCP			0.50	(0.10, 0.88)	-0.13	(-0.47, 0.21)	0.20	(0.03, 0.37)
CAM-S Peak:								
0-2	57.61	(57.12, 58.09)	-0.64	(-1.04, -0.24)	1.12	(0.77, 1.47)	-0.09	(-0.26, 0.07)
3-7	57.61	(57.12, 58.09)	-0.80	(-1.22, -0.39)	1.35	(0.98, 1.72)	-0.26	(-0.44, -0.06)
8-19	57.61	(57.12, 58.09)	-2.30	(-3.24, -1.36)	2.57	(1.86, 3.28)	-0.98	(-1.38, -0.57)
Sensitivity analysis: worst case scenario								
Baseline GCP			0.50	(0.10, 0.88)	-0.14	(-0.48, 0.20)	0.26	(0.09, 0.43)
CAM-S Peak:								
0-2	57.61	(57.12, 58.09)	-0.64	(-1.04, -0.24)	1.17	(0.83, 1.52)	-0.28	(-0.45, -0.12)
3-7	57.61	(57.12, 58.09)	-0.80	(-1.22, -0.39)	1.38	(1.01, 1.75)	-0.38	(-0.58, -0.19)
8-19	57.61	(57.12, 58.09)	-2.30	(-3.25, -1.36)	2.49	(1.78, 3.20)	-0.69	(-1.09, -0.28)

GCP = general cognitive performance, CAM-S = Confusion Assessment Method-Severity score, CI = confidence interval

Interpretation of Results for Sensitivity Analysis 1

For the GCP analysis, in both the best and worst case scenario, we arrive at comparable decisions regarding the effect of delirium severity in long term slope (last column in Table S3). In the observed data, best case, and worst case scenario the slope is declining among all severity groups over the 36 month interval. In the best case scenario (i.e., those persons who dropped out were on a steeper cognitive decline trajectory than predicted by their observed data) the dose-response effect is more pronounced than in the observed data. In the worst case scenario (i.e., persons who dropped out were on a much lower decline in cognitive functioning relative to what would be inferred from their observed data), the dose-response effect is more subtle. These differences in the patterns of results set boundaries on the range of plausible effects of delirium in our study. The original maximum likelihood results reported in the top segment, and in the main manuscript, reflect our best estimate of the population parameters, and we believe that if the missing data mechanism is not MAR, the results would be somewhere between the observed results and those of the best case scenario. That is to say, it is more plausible that persons with more severe delirium who went on to die or leave the study due to institutionalization would have steeper cognitive decline slopes than what might be expected given their observed data, rather than shallower cognitive decline slopes. Therefore, we believe that if anything our maximum likelihood results are accurate or perhaps somewhat conservative estimates of the true population parameters.

Sensitivity Analysis 2: Assessing the Relationship between Sum of CAM-S Scores and Long-Term Cognitive Decline

Table S4. Relationship of Peak Confusion Assessment Method-Severity (CAM-S) to Sum of all CAM-S Scores

Sum CAM-S score	Peak CAM-S score			Total
	0-2	3-7	8-19	
0-2	112	0	0	112
3-16	131	222	14	367
17-max	1	26	54	81
Total	244	248	68	560

Table S5. Sum of CAM-S Scores by Corrected GCP Scores over Time

Visit	Full Sample		Sum CAM-S score					
			0 – 2		3 – 16		17 – max	
Month	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
0	560	57.6 (7.3)	112	60.9 (6.2)	367	57.6 (7.3)	81	53.3 (6.1)
1	548	56.8 (7.9)	112	60.3 (6.2)	359	57.1 (7.8)	77	50.6 (6.5)
2	536	58.0 (7.9)	109	61.4 (6.3)	352	58.1 (8.0)	75	52.6 (6.2)
6	528	58.2 (7.5)	109	61.4 (6.1)	347	58.2 (7.6)	72	53.4 (6.9)
12	511	58.4 (7.6)	105	61.7 (6.6)	334	58.5 (7.6)	72	53.3 (6.3)
18	499	58.3 (8.0)	104	61.5 (6.7)	323	58.6 (7.7)	72	52.2 (7.7)
24	474	58.2 (8.0)	101	61.3 (6.2)	311	58.4 (7.9)	62	51.9 (7.9)
30	325	57.5 (8.2)	67	61.1 (7.2)	213	57.3 (8.2)	45	52.8 (7.4)
36	312	57.1 (8.4)	62	61.4 (7.3)	201	57.1 (8.2)	49	52.0 (7.8)

CAM-S = Confusion Assessment Method-Severity

GCP = General Cognitive Performance

Table S6. Empirically Observed Prevalence of Proxy IQCODE Impairment over Time

Visit month	Full Sample		Sum CAM-S score					
	N	<i>n</i> (%)	0 – 2		3 – 16		17 – max	
			N	<i>n</i> (%)	N	<i>n</i> (%)	N	<i>n</i> (%)
0	548	118 (22)	111	20 (18)	356	71 (20)	81	27 (33)
6	514	135 (26)	107	17 (16)	336	91 (27)	71	27 (38)
12	487	130 (27)	102	15 (15)	320	89 (28)	65	26 (40)
18	480	142 (30)	99	12 (12)	311	96 (31)	70	34 (49)
24	452	125 (28)	97	20 (21)	296	82 (28)	59	23 (39)
30	314	101 (32)	65	10 (15)	203	67 (33)	46	24 (52)
36	287	94 (33)	59	10 (17)	184	61 (33)	44	23 (52)

CAM-S = Confusion Assessment Method-Severity

IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly

N=total possible sample, n=number with proxy IQCODE impairment

Figure S2. Trajectory of Estimated General Cognitive Performance by Sum of Confusion Assessment Method-Severity (CAM-S) Groups

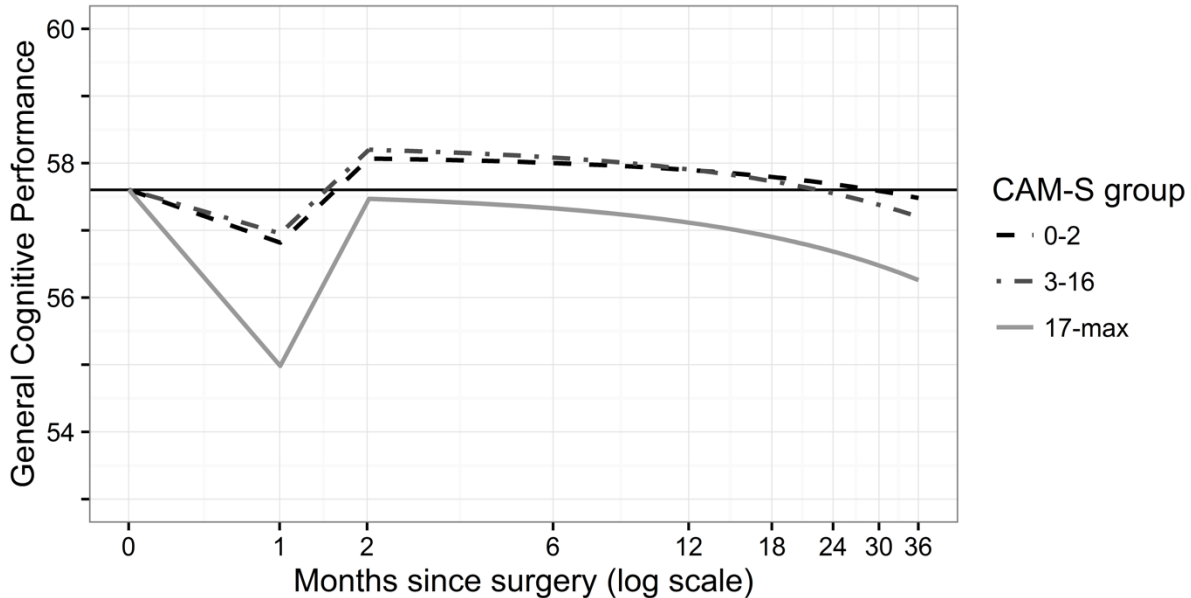


Figure S2 Legend:

Figure S2 demonstrates the relationship between estimated general cognitive performance (GCP) and time following surgery (months, natural log scale) by delirium severity groups. The model is adjusted for baseline GCP, age, gender, non-white race, education, Charlson score, Geriatric Depression score, instrumental activities of daily living (IADL) impairment, surgery type, and proxy Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) impairment. The solid horizontal line indicates the average GCP score prior to surgery. Delirium severity group is observed after baseline, and is dependent on baseline, and therefore the figure is plotted showing no mean difference in GCP at baseline by delirium severity group. The amount of punctuation (acute decline at one month), recovery (up to two months), and long-term decline (two to 36 months) is shown by each CAM-S severity group, 0-2 (dashed black line), 3-16 (dot-dashed black line) and 17-max (solid grey line). In the acute (punctuation) phase, all groups decline with the most severe group declining the most. This is followed by recovery of cognitive performance, with the less severe groups recovering (at two months) past their baseline (0 months) GCP score, and those in the most severe group showing an incomplete return to baseline. Over long-term follow-up, the less severe groups gradually decline in GCP performance, whereas the most severe group demonstrates a more accelerated rate of decline.

Figure S3. Predicted Prevalence of Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) Impairment by delirium severity group (sum CAM-S) and study month

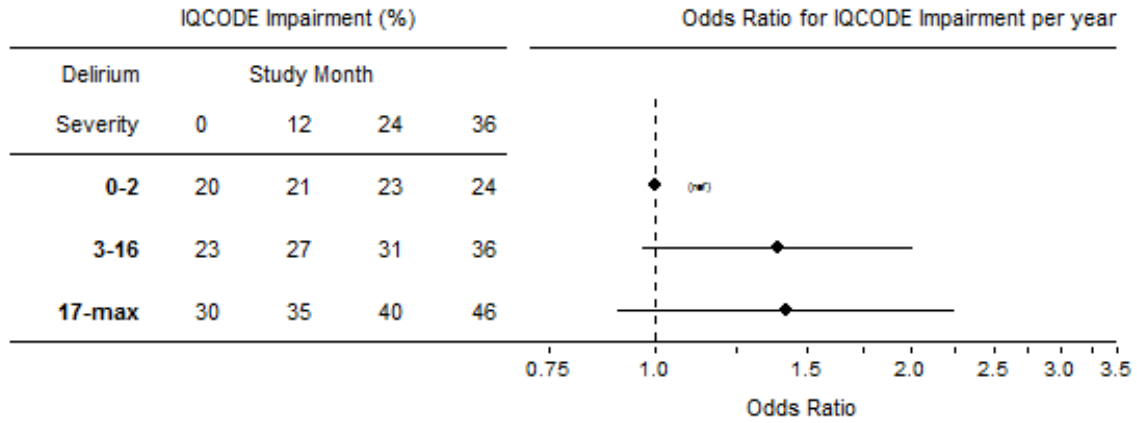


Figure S3 legend:

Figure S3 demonstrates the relationship between the prevalence of proxy IQCODE impairment (score ≥ 3.2) over time (study month) following surgery, calculated using a mixed effects generalized linear model. The model controlled for: age, gender, non-white race, education, Charlson score, Geriatric Depression Scale score, and surgery type. Model-implied (or expected) proportions with IQCODE impairment given mean values on covariates are presented in the table. Odds ratios (and 95% confidence bands) illustrate the size and precision of estimates of the delirium severity group by time (in years following surgery) interaction effects. Over time, all groups have increasing probability of being classified as impaired on the IQCODE, but the slope over time is significantly ($p=.02$) faster only for patients with the most severe delirium (sum of CAM-S scores 17-max). The per-year odds of IQCODE ≥ 3.2 for this group is about 2.0 times greater than that observed for the lowest delirium severity group.

Table S7. Detailed Model Results Reporting the Association between Delirium Severity (Sum Confusion Assessment Method-Severity) and General Cognitive Performance (N=560)

Model Parameter	Estimate	95% CI	P-value
Effect of baseline GCP in ...			
GCP punctuation	0.44	(0.01, 0.85)	.043
GCP recovery	-0.12	(-0.48, 0.23)	.497
GCP slope, Months 2 - 36	0.27	(0.08, 0.46)	.006
Estimated means for latent growth model effects by Delirium (sum CAM-S) severity group			
CAM-S Sum 0-2			
GCP Punctuation	-0.74	(-1.33, -0.13)	.017
GCP Recovery	1.24	(0.70, 1.78)	<.001
GCP Slope, Months 2 - 36	-0.18	(-0.42, 0.07)	.167
CAM-S Sum 3-16			
GCP Punctuation	-0.60	(-0.94, -0.25)	.001
GCP Recovery	1.23	(0.93, 1.53)	<.001
GCP Slope, Months 2 - 36	-0.32	(-0.49, -0.15)	<.001
CAM-S Sum 17-max			
GCP Punctuation	-2.57	(-3.38, -1.75)	<.001
GCP Recovery	2.47	(1.77, 3.17)	<.001
GCP Slope, Months 2 - 36	-0.39	(-0.81, 0.02)	.065

CAM-S = Confusion Assessment Method-Severity score, GCP = general cognitive performance

Figure S4. STROBE Diagram of the Analytic Sample

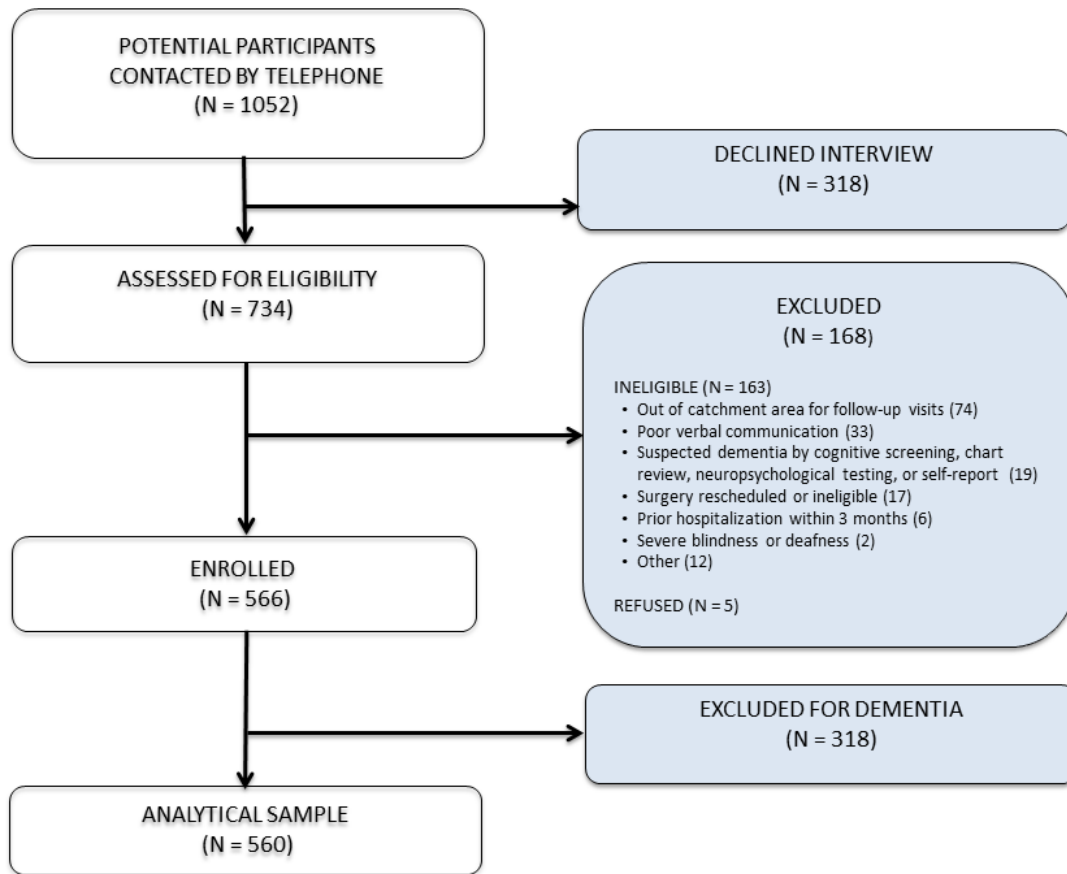


Table S8. SAGES Follow-up Success Rate

	Potential Interviews	Pending Interviews	Completed Interview N (%)	Refused or Unobtainable	Deaths	Drop-Outs
1 Month	560	0	550 (98)	2	1	7
2 Month	552	0	537 (97)	10	1	4
6 Month	547	0	531 (97)	10	3	3
12 Month	541	0	513 (95)	23	3	4
18 Month	534	0	504 (94)	26	3	3
24 Month	520	8*	478 (92)	19	10	13
30 Month	360	145*	326 (91)	25	7	2
36 Month	321	175*	316 (98)	2	3	0

*Pending interviews indicates that the subjects have not yet reached the time for their scheduled follow-up interviews.

Table S9. Description of Study Sample by Death or Nursing Home Placement Status

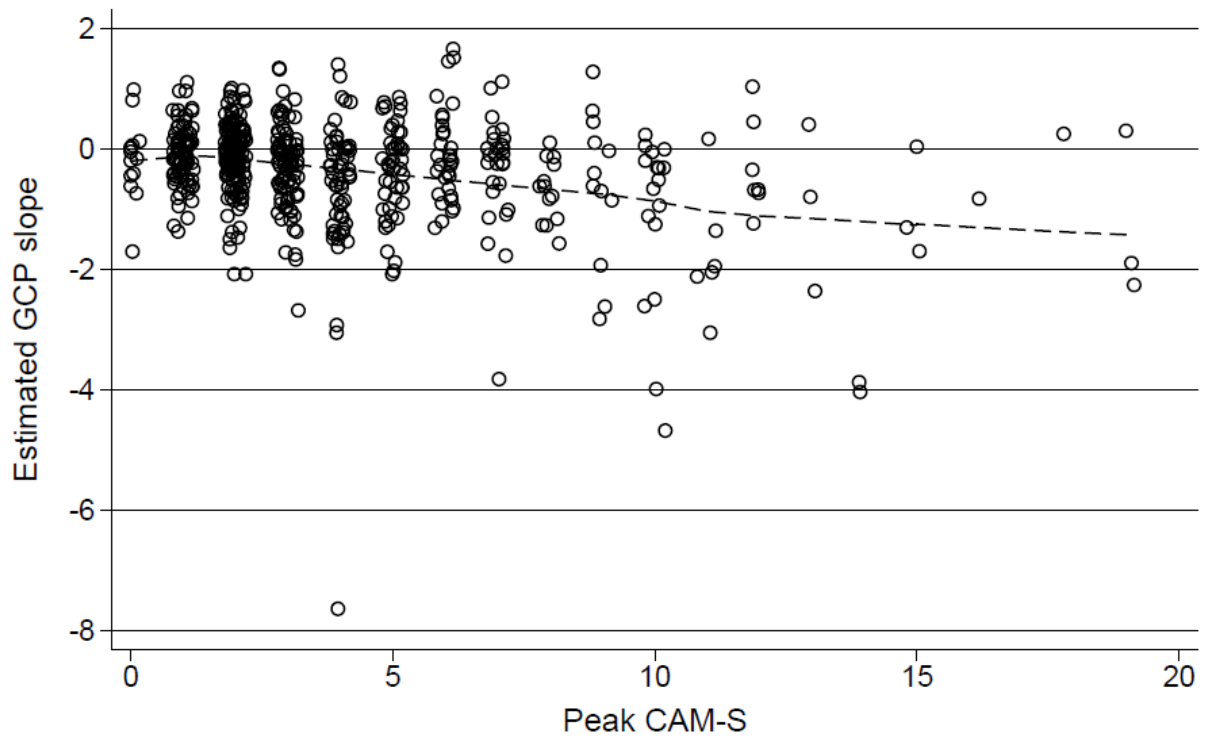
Characteristic	Full Sample (N = 560)	Death or Nursing Home Placement (N = 103)	No Death or Nursing Home Placement (N = 457)
Age - mean (SD)	76.7 (5.2)	78.9 (5.6)	76.2 (4.9)
Female – n (%)	326 (58)	62 (60)	264 (58)
Nonwhite – n (%)	42 (8)	6 (6)	36 (8)
Education – mean years (SD)	15.0 (2.9)	15.1 (3.0)	14.9 (2.9)
Married – n (%)	332 (59)	52 (50)	280 (61)
Lives Alone – n (%)	167 (30)	36 (35)	131 (29)
Charlson score - n (%)			
0	257 (46)	36 (25)	221 (48)
1	139 (25)	23 (22)	116 (25)
2+	164 (29)	44 (43)	120 (26)
GDS15 score - n (%)			
0 - 5	489 (88)	84 (82)	405 (89)
6 - 15	69 (12)	19 (18)	50 (11)
GCP score - mean (SD)	57.6 (7.3)	55.5 (7.6)	58.1 (7.1)
3MS score - n (%)			
85-100	523 (93)	96 (93)	427 (93)
71-84	37 (7)	7 (7)	30 (7)
Proxy IQCODE (baseline) - n (%)			
Not Impaired	430 (78)	74 (72)	356 (78)
Impaired	118 (22)	26 (26)	92 (21)
ADL impairment – n (%)	42 (8)	14 (14)	28 (6)
IADL impairment – n (%)	152 (27)	40 (39)	112 (25)
Surgery type - n (%)			
Orthopedic	454 (81)	81 (79)	373 (82)
Vascular	35 (6)	8 (8)	27 (6)
General	71 (13)	14 (14)	57 (12)
Delirium ^a - n (%)			
None	426 (76)	74 (72)	352 (77)
Delirium	134 (24)	29 (28)	105 (23)
Peak CAM-S Score	1.7 (0.7)	1.8 (0.7)	1.7 (0.7)
Peak CAM-S Score – n (%)			
0 – 2	244 (44)	37 (36)	207 (45)
3 – 7	248 (44)	48 (47)	200 (44)
8 – 19	68 (12)	18 (17)	50 (11)

^aDelirium status was determined with daily interviews rating the Confusion Assessment Method, augmented by a validated chart review

ADL = Activities of Daily Living, impairment indicated by human assistance to complete any activity
CAM-S = Confusion Assessment Method-Severity

GCP = General Cognitive Performance, composite measure of neuropsychological measures reflecting cognitive domains vulnerable to delirium, see text for details
GDS15= Geriatric Depression Scale 15 point version, range (0-15), higher is worse; a score 6 and above is considered impaired
IADL = Instrumental Activities of Daily Living, impairment indicated by human assistance to complete any activity
IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly, score >3.2 indicates cognitive impairment
3MS = Modified Mini-Mental State Exam, range (0-100), lower score indicates impairment; a score ≤ 84 is considered impaired
SAGES = Successful Aging after Elective Surgery Study
SD= standard deviation.
The Charlson comorbidity score ranged from 0-35, with higher scores indicating more comorbidity.

Figure S5. Continuous Peak CAM-S Measure by GCP slope



* Estimated GCP slope is the estimated linear annual change in GCP from month 2 to 36.

Matched Analysis

Upon the excellent suggestion of an anonymous reviewer, we tested our hypothesis that delirium was related to cognitive decline using a matched design. Surgical patients in the most severe post-operative delirium category (peak CAM-S 8-19, N=68) were matched on the basis of pre-operative GCP score coarsened within 0.4 population SD units to patients in the peak CAM-S 3-7 group and patients in the peak CAM-S 0-2 group (mildest group). We used coarsened exact matching⁴ to match patients in the most severe group (peak CAM-S 8-19) with at least two observations in the 2-36 month follow-up window to one patient from the peak CAM-S 3-7 score group and another patient from the peak CAM-S 0-2 group. Matching was performed on the basis of pre-operative GCP score binned in units of 4 (0.4 population SD units). We were able to match 62 of 68 patients. The mean (SD) GCP score at baseline had a mean of 54 and the standard deviation was 5 in each of the three groups. The pace of decline was faster in the peak CAM-S 8-19 group (slope -0.090 population SD units per year) than in the peak CAM-S 3-7 group (slope -0.035 SD/year) which was faster than the peak CAM-S 0-2 group (slope -.017 SD/year). The difference between the 0-2 group and 8-19 group was statistically significant (difference of .073 SD/year, P = .01) but was not different between the 0-2 group and the 3-7 group (difference of .018 SD/year, P = .55). These findings are consistent with the findings we report in the manuscript using the covariate adjusted but not matched patients.

Table S10. Results of Matched Analysis (N=186, n=62 in each delirium subgroup)

	Peak CAM-S Group		
	0-2 (n=62)	3-7 (n = 62)	8-19 (n = 62)
Baseline GCP, mean (SD)	54.2 (5.0)	54.2 (5.2)	54.2 (5.0)
Mixed model results			
Difference at baseline (GCP, est, P)	--	-1.1 (.31)	-1.3 (.22)
Slope from M2-M36 (GCP/y; est, P)	-0.17 (.44)	--	--
Difference in slope (GCP/y; est P)	--	-0.18 (.55)	-.73 (.01)

CAM-S=Confusion Assessment Method-Severity; GCP=general cognitive performance; M=month; SD=standard deviation

References

1. Jones RN. Practice and retest effects in longitudinal studies of cognitive functioning. *Alzheimers Dement.* 2015;1(1):101-102.
2. Lewis M, Maruff P, Silbert B. Statistical and conceptual issues in defining post-operative cognitive dysfunction. *Neurosci. Biobehav Rev.* 2004;28(4):433-440.
3. Raftery AE. Bayesian model selection in social research. *Soc Method.* 1995;25:111-164.
4. Iacus SM, King G, Porro G. Causal inference without balance checking: Coarsened exact matching. *Political Analysis.* 2012;20(1):1-24.