Supplemental Online Content

Kunicki ZJ, Ngo LH, Marcantonio ER, et al. Six-year cognitive trajectory in older adults following major surgery and delirium. *JAMA Intern Med*. Published online March 20, 2023. doi:10.1001/jamainternmed.2023.0144

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

eTable 1. Participant flow and follow-up through 72-month follow-up.

Time point	Number in sample at assessment time point	Dropout in interval	Died in interval	In sample at end of interval	Number assessed at time point
Overall					
Baseline	560	0	0	560	560
1 month	560	7	1	552	548
2 months	552	6	2	544	536
6 months	544	2	4	538	528
12 months	538	8	3	527	511
18 months	527	7	6	514	499
24 months	514	12	9	493	474
30 months	493	6	11	476	325
36 months	476	4	3	469	456
48 months	469	32	24	413	385
60 months	413	13	21	379	354
72 months	379	18	26	335	313
Delirium					
Baseline	134	0	0	134	134
1 month	134	2	0	132	129
2 months	132	2	0	13	126
6 months	130	0	1	129	124
12 months	129	2	1	126	120
18 months	126	2	1	123	117
24 months	123	4	3	116	106
30 months	116	1	4	111	83
36 months	111	0	0	111	110
48 months	111	4	9	98	90
60 months	98	4	9	85	79
72 months	85	3	11	71	65
No Delirium					
Baseline	426	0	0	426	426
1 month	426	5	1	420	419
2 months	420	4	1	415	410
6 months	415	2	4	409	404
12 months	409	6	2	401	391
18 months	401	5	5	391	382
24 months	391	8	6	377	368
30 months	377	5	7	365	242
36 months	365	4	3	358	346
48 months	358	28	15	315	295
60 months	315	9	12	294	275
72 months	294	15	15	264	248

eAppendix 1 - Statistical modeling

Overview: This appendix describes the modeling steps we pursued that culminated in our estimation of a random effects piecewise linear trajectory model (our components of change model). We describe the multiple models we examined and contrast with the components of change model.

Procedures: Our analysis proceeded by first identifying a reasonable trajectory for the repeated cognitive measures (GCP) from preoperative baseline to 72-month follow-up, followed by testing our hypothesis that change differed according to post-operative delirium status. In our initial exploratory step, we used multiple methods: mixed effect regression models, random effects models, and random changepoint models.

1.0 Random changepoint model

We initiated our longitudinal data analysis approach to the SAGES repeated cognitive data (GCP, general cognitive performance) using a random changepoint model. Because of the observed acute decline and immediate recovery effects in the SAGES sample,[1] we used data from month 6 to month 72 in the random change point model. This decision was made so the estimated change point did not capture either the punctuation or recovery effect, but rather change in the trajectory observed after 6 months in previous studies.[1] We prepended a latent difference score model[2] onto the latent difference score model to capture the change from baseline to month 1, and month 1 to month 2, before the estimating the random change point model from months 6 to 72. Covariances between the latent difference score and random change point models were estimated so each time point was taken into account in both models. We also regressed delirium and the other demographic covariables from the components of change model onto each latent variable.

Results: The results of the random changepoint model are shown in **Table e2**. The random change point occurred 1.6 years after surgery for the non-delirium group and 1.7 years after surgery for the delirium group. The mean of the slope in the no delirium group prior to the change point implied continuing improvement in GCP scores (+0.3 GCP points/month), while the mean of the slope prior to the change point implied net cognitive decline in the delirium group (-0.2 GCP points/month).

Conclusion: Because of the qualitative mean difference in pre change-point slopes, the meaning of the changepoint is ambiguous and possibly different in the delirious and non-delirious samples. In the non-delirious, the change point could signify when cognitive recovery ends. In the delirious, the change point could signify an

estimate of a point when cognitive change accelerates. We felt that this ambiguity warranted exploring the data with additional models.

eTable 2. Random change point model results.

Coefficient	No Delirium	Delirium	Difference
Intercept	57.6	57.6	0.0
Punctuation	-0.7	-1.7	-1.0
Recovery	0.8	1.8	1.0
6-Month intercept	57.9	58.0	0.1
Pre-change slope	0.3	-0.2	-0.4
Post-change slope	-1.0	-1.5	-0.5
Random change point (in years from baseline)	1.6	1.7	0.1

Note: Table entries are parameter estimates from random changepoint model

2.0 Random coefficient models with linear splines

We followed on our random changepoint models with random coefficient models, building upon a modeling approach we have used in our report of cognitive trajectories following elective surgery and delirium through 36 months in the same cohort.[1] We build a model that includes multiple piecewise linear splines with knots at months 1, and 2 to model the observed acute decline, recovery, and long-term trajectory. We also constructed alternative functional forms, including linear, quadratic (also with knots at months 1 and 2) and piecewise linear with a knots at: months 1, 2, and 12; months 1, 2, and 18; and months 1, 2, 6, and 18. We also estimated models with random change points.[3] Akaike information criteria (AIC) and Bayesian information criteria (BIC) for each model was compared, and the model with the lowest AIC and BIC was considered the best fitting model. AICs and BICs for each model are shown in **Table e3**.

eTable 3. Model fit summaries for alternative functional form models

Models	Free Parameters	Loglikelihood	AIC	BIC
Linear	17	-14342	28717	28791
Linear segmented 1.2	19	-14219	28476	28558
Quadratic	21	-14196	28435	28526
Linear segmented 1.2.12	23	-14159	28365	28464
Linear segmented 1.2.18	23	-14150	28346	28445
Linear segmented 1.2.6.18	28	-14120	28296	28417
Random changepoint	42	-14089	28262	28444
Components of change	25	-14145	28309	28403

Notes: 1.2, knots at months 1 and 2; 1.2.12, knots at months 1, 2, and 12; 1.2.18, knots at months 1, 2, and 18, 1.2.6.18, knots at months 1, 2, 6, and 18.

The best fitting model included fixed effects at months 1 and 2 and knots at both month 6 and 18. Examination of the model trajectory suggested there was a positive slope (Est. = 0.33, p = .01) between months 6 and 18. This implied there was continued recovery up to 1.5 years post-surgery. We sought to better explore this longer-term recovery, and the impact delirium may have on longer-term recovery, which led to the components of change model.

3.0 Components of change model

The components of change model, is a random effects model comprised of five latent variables. The design matrix for the components of change model was:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & t_1 & 1 & 0 & 0 \\ 1 & t_2 & 1 & 1 & 0 \\ 1 & t_6 & 1 & 1 & * \\ 1 & t_{12} & 1 & 1 & * \\ 1 & t_{18} & 1 & 1 & * \\ 1 & t_{24} & 1 & 1 & * \\ 1 & t_{30} & 1 & 1 & 1 \\ 1 & t_{36} & 1 & 1 & 1 \\ 1 & t_{48} & 1 & 1 & 1 \\ 1 & t_{60} & 1 & 1 & 1 \\ 1 & t_{72} & 1 & 1 & 1 \end{bmatrix}$$

Where the first column represented the intercept, the second the linear slope (i.e., “long-term period”), the third the fixed effect (i.e., “acute period”) at month 1, the fourth the fixed effect (i.e., “post-acute period”) at month 2,

and the final column represent the second slope modeling the intermediate-period. The * for the long-term recovery slope indicate the loading was freely estimated. The estimated loadings followed a Gompertz function, which indicates slower growth at the start of a time period and faster growth near the end of the time period. We used the estimated Gompertz function based on the freely estimated time steps to generate fixed time steps, which we used for analysis. The design matrix including the fixed time steps, estimated from the Gompertz function, was:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & t_1 & 1 & 0 & 0 \\ 1 & t_2 & 1 & 1 & 0 \\ 1 & t_6 & 1 & 1 & .20 \\ 1 & t_{12} & 1 & 1 & .51 \\ 1 & t_{18} & 1 & 1 & .79 \\ 1 & t_{24} & 1 & 1 & .96 \\ 1 & t_{30} & 1 & 1 & 1 \\ 1 & t_{36} & 1 & 1 & 1 \\ 1 & t_{48} & 1 & 1 & 1 \\ 1 & t_{60} & 1 & 1 & 1 \\ 1 & t_{72} & 1 & 1 & 1 \end{bmatrix}$$

In **Table e4** below, we provide detailed models results which expand upon Table 3 in the main manuscript. These results show the effect of the included model covariables (age, gender, non-White race, education, Charlson score, GDS score, IQCODE, impairment in instrumental activities, and surgery type) on each of the latent variables in the components of change model.

eTable 4. Effect of covariables on component of change model

Model term	Estimate	SE	p-value
Intercept regressed on			
Age at surgery	-0.42	0.05	<.001
Non-White or Hispanic	-7.85	1.06	<.001
Female sex	0.98	0.54	.07
IQCODE	-0.38	1.23	.76
Any IADL impairment	-1.84	0.7	.01
Geriatric Depression Scale score	-0.41	0.12	<.001
Charlson score	-1.06	0.31	<.001
Vascular surgery	-1.59	1.06	.13
Gastrointestinal surgery	0.94	0.74	.21
Acute period regressed on			
Delirium	-1.18	0.34	<.001
Age at surgery	-0.01	0.03	.79
Non-White or Hispanic	0.20	0.52	.70
Female sex	0.56	0.28	.04
IQCODE	-0.21	0.58	.71
Any IADL impairment	-0.75	0.34	.03
Geriatric Depression Scale score	0.00	0.06	.94
Charlson score	-0.04	0.17	.80
Vascular surgery	-1.15	0.62	.07
Gastrointestinal surgery	0.66	0.36	.07
Post-acute period regressed on			
Delirium	1.12	0.32	<.001
Age at surgery	0.01	0.03	.82
Non-White or Hispanic	-0.47	0.49	.33
Female sex	-0.3	0.26	.25
IQCODE	-0.65	0.55	.24
Any IADL impairment	0.56	0.31	.07
Geriatric Depression Scale score	0.03	0.05	.62
Charlson score	-0.12	0.15	.43
Vascular surgery	0.13	0.57	.81
Gastrointestinal surgery	-1.05	0.36	.004
Intermediate period regressed on			
Delirium	0.10	0.47	.82
Age at surgery	0.10	0.04	.03
Non-White or Hispanic	-0.25	0.68	.71
Female sex	0.53	0.38	.16
IQCODE	-0.14	0.77	.85
Any IADL impairment	-0.65	0.55	.24
Geriatric Depression Scale score	-0.16	0.09	.05
Charlson score	0.38	0.25	.13
Vascular surgery	0.80	1.00	.42
Gastrointestinal surgery	-0.26	0.54	.63

Model term	Estimate	SE	p-value
Long-term period regressed on			
Delirium	-0.39	0.16	.01
Age at surgery	-0.07	0.01	< 001
Non-White or Hispanic	0.36	0.18	.05
Female sex	-0.14	0.12	.24
IQCODE	-0.23	0.29	.41
Any IADL Impairment	0.04	0.16	.80
Geriatric Depression Scale score	0.04	0.02	.09
Charlson score	-0.17	0.07	.02
Vascular surgery	-0.05	0.28	.85
Gastrointestinal surgery	0.05	0.18	.77
Delirium† regressed on			
Age at surgery (per year)	-0.004	0.02	.86
Non-White or Hispanic	-0.19	0.40	.64
Female sex	0.15	0.22	.50
IQCODE	1.23	0.42	.003
Any IADL impairment	0.12	0.24	.60
Geriatric Depression Scale score	0.02	0.04	.57
Charlson score	0.27	0.13	.04
Baseline GCP (*Intercept*, per GCP point)	-0.06	0.02	<.001
Vascular surgery	0.05	0.41	.90
Gastrointestinal surgery	0.19	0.32	.56

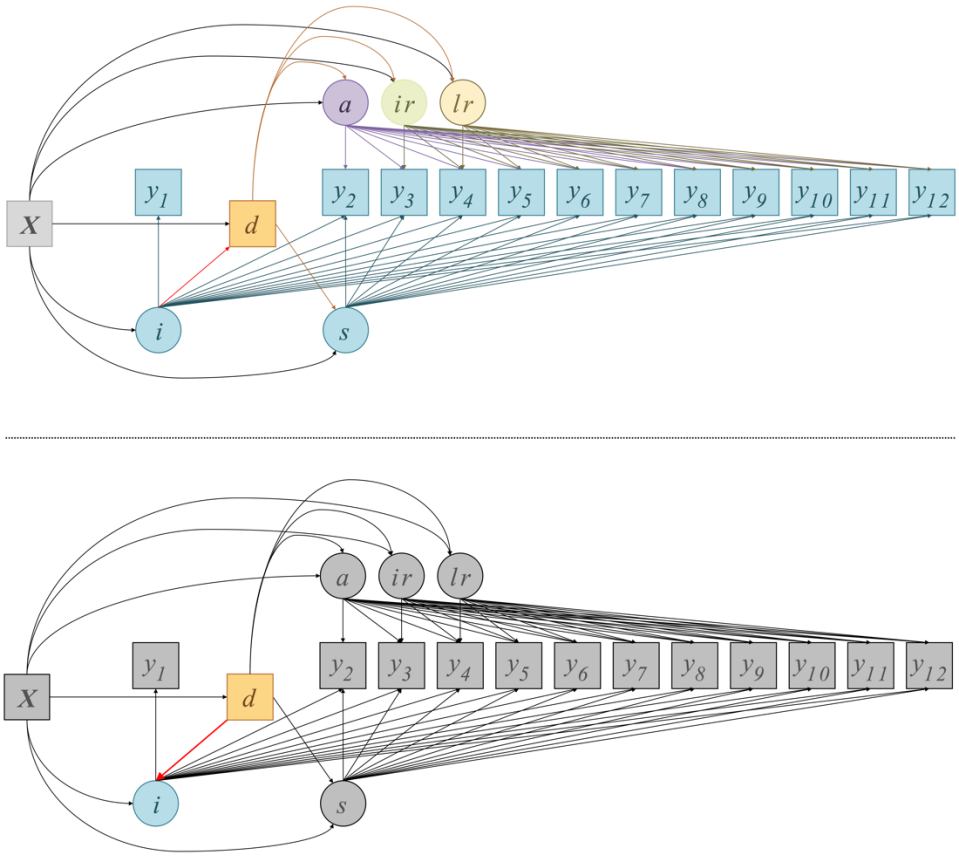
Model term	Estimate	SE	p-value
Latent variable intercepts			
Intercept (baseline)	57.60	0.26	.003
Acute period	-0.46	0.16	< .001
Post-acute period	1.10	0.14	< .001
Intermediate period	1.89	0.22	< .001
Long-term period	-1.01	0.07	< .001
Latent variable residual variances			
Intercept (baseline)	37.78	2.57	< .001
Acute period	-		
Post-acute period	-		
Intermediate period	4.00	1.39	.004
Long-term period	0.80	0.19	< .001
Latent variable covariances			

Intercept (baseline) with *Long-term period*	0.92	0.28	.001
Intermediate period with *Long-term period*	-0.88	0.39	.02

Notes: IQCODE, Informant Questionnaire on Cognitive Decline in the Elderly; IADL, Instrumental Activities of Daily Living, SE, standard error. † Parameter estimates for the regression of delirium on covariables are logistic regression coefficients. All other table entries labeled regression parameters are linear regression coefficients

3.1 Components of change model, delirium as covariable

We also estimate a form of the components of change model where delirium is included as a covariable. Our main model, appropriately specified with respect to the temporal ordering of baseline GCP and delirium risk, regresses delirium on baseline GCP. The delirium as covariable model regresses baseline GCP on delirium group. This analogous to using a two-sample t test to evaluate the mean difference in baseline GCP score according to whether or not delirium occurred in the hospital stay. We use this model as a convenience tool to easily obtain delirium group baseline mean differences in GCP. When results pertain to the delirium as a covariable model, those are clearly identified in this appendix and the main text. The figure below (**Figure e1**) illustrates the difference between the delirium as an intermediate outcome (top panel) and delirium as a covariable (bottom panel) models.



eFigure. Delirium as an intermediate outcome and delirium as a covariable components of change model. The top panel corresponds to our main model. The bottom panel highlights the difference in the delirium as a covariable model. These path diagrams illustrate how repeatedly observed GCP scores from baseline

through month 72 (y1-y12, corresponding to observations at post-surgical month 0, 1, 2, 6, 12, 18, 24, 30, 36, 48, 60, 72) and observed post-operative delirium (*d*) are related to intercept (*i*) and slope (*s*) growth functions, and acute decline, immediate recovery, and longer-term recovery (*a*, *ir*, *lr*) effects. ***X*** denote the multiple covariables (age, sex, race/ethnicity, IQCODE score, IADL impairment, Charlson comorbidity, depression, and surgery type). For the sake of clarity, residual variances for all *y*'s, latent variables (*i*, *s*, *a*, *ir*, *lr*) are omitted, as are covariances among the latent variables.

eAppendix 2 – Exploring the effects of non-normality in GCP scores in the estimation of long-term cognitive aging slopes

Objective: The goal of this analysis was to consider if non-normality in the distribution of GCP scores could be influencing the estimated slopes in aging-related cognitive decline (the slope parameter in the components of change model). This was motivated by our finding that the long-term slope estimates in the current analysis of cognitive change over 72-months was considerably larger than our estimate of the long-term slope in our analysis of data from this cohort followed over 36 months and published previously.[1]

Procedure: We repeated our components of change model using a transformation of the GCP score. The transformation was a rank-based normalization transformation, known as the Blom transformation.[4] Essentially, the Blom transformation involves converting the scores to ranks, and then converting the ranks to percentile ranks, and applying an inverse normal transformation to the percentile values. To increase the comparability of Blom-transformed GCP to our original metric GCP, we linearly scale the Blom-transformed values using the original metric GCP mean and standard deviation.

Results: After the Blom transformation, we fit the components of change model with covariable adjustment for age, race/ethnicity, sex, IQCODE, IADLS, GDS, and surgery type to the data. We display the results of the original components of change model and Blom-transformed components of change model in **Table e5**. The results suggest there are minor differences, at the tenth or hundredth decimal place, between the original and Blom-transformed data. These minor differences are seen in both the estimates of the intercept, acute period, post-acute period, intermediate period, and Long-term period effects as well as the estimates of delirium on each of these five variables. Based on these results, sensitivity to departures from normality do not impact the results of the components of change model in the presence or absence of delirium.

eTable 5. Parameter estimates from components of change model using original GCP scaling and Blom-transformed GCP

	Original GCP scaling				Blom-transformed GCP			
Coefficient	No Delirium	Delirium	Difference	SE	No Delirium	Delirium	Difference	SE
Intercept	58.2	55.8	-2.4	0.6	58.2	55.7	-2.5	0.6
Acute period	-0.5	-1.7	-1.2	0.3	-0.4	-1.5	-1.1	0.3

Post-acute period	1.1	2.2	1.1	0.3	1.1	2.2	1.0	0.3
Intermediate period	1.9	2.0	0.1	0.5	1.9	1.8	-0.1	0.4
Long-term cognitive decline	-1.0	-1.4	-0.4	0.1	-0.9	-1.2	-0.3	0.1

Note: The table entries are estimated from a purposefully time-order mis-specified model (with respect to the effect of baseline GCP in delirium risk), to more easily evaluate the baseline differences in GCP score according to delirium group.

eAppendix 3 – Exploring the effect of different missing data handling strategies on main results

Objective: The goal of the sensitivity analyses described in this appendix was to explore the impact of different procedures for handling missing data, and the extent to which missing data may have on the substantive conclusions from our models. We compare our preferred approach (maximum likelihood estimation) to multiple imputation, extreme value imputation, and complete case analysis.

Study 1: Maximum Likelihood vs Multiple Imputation

Procedure: In our first missing data sensitivity analysis, we used multiple imputation (MI) with age, race/ethnicity, sex, IQCODE, IADLS, GDS, and surgery type as auxiliary variables in the imputation routine. We display the results of the original components of change model and multiple imputation model in **Table e6**.

eTable 6. Parameter estimates from components of change model using maximum likelihood estimation and multiple imputation

	Maximum Likelihood Estimates				Multiple Imputation Estimates			
Coefficient	No Delirium	Delirium	Differ-ence	SE	No Delirium	Delirium	Differ-ence	SE
Intercept	58.2	55.8	-2.4	0.6	58.2	55.8	-2.5	0.6
Acute period	-0.5	-1.7	-1.2	0.3	-0.5	-1.7	-1.2	0.3
Post-acute period	1.1	2.2	1.1	0.3	1.2	2.2	1.0	0.3
Intermediate period	1.9	2.0	0.1	0.5	1.8	1.7	-0.1	0.4
Long-term cognitive decline	-1.0	-1.4	-0.4	0.1	-1.0	-1.3	-0.3	0.1

Note: The table entries are estimated from a purposefully time-order mis-specified model (with respect to the effect of baseline GCP in delirium risk), to more easily evaluate the baseline differences in GCP score according to delirium group.

Results: The results of **Table e6** suggest there are minor differences, at the tenth decimal place, between the ML and MI imputed data results. These minor differences are seen only in the recovery and longer-term recovery terms, and in the effect of delirium on recovery, longer-term recovery, and rate of cognitive decline (i.e., slope).

Study 2: Maximum Likelihood vs Extreme Value Imputation

Procedure: The second missing data sensitivity analysis involved imputing extreme values for participants with missing follow-up data. Extreme values were defined as the 97.5th percentile of the expected GCP value by time point for participants who experienced postoperative delirium, and the 2.5th percentile of the expected GCP value by time point for participants who did not experience postoperative delirium. In other words, we imputed the best-case scenario GCP values for participants who experienced delirium, and the worst-case scenario values for participants who did not experience delirium. These best-case and worse-case descriptions are in reference to our hypothesis that delirium is associated with a faster pace of cognitive aging.

Results: The results of the extreme value imputation, shown in **Table e7**, suggest the slope for the no delirium group was less steep for extreme value imputation compared to maximum likelihood imputation (-0.5 vs -1.0, respectively). The effect of delirium was also lessened (-0.3 vs -0.4 for extreme value vs maximum likelihood,

194 respectively). However, we still observed the same pattern of results as shown in the maximum likelihood
195 imputation approach.
196

eTable 7. Parameter estimates from components of change model using maximum likelihood estimation and extreme value imputation

	Maximum Likelihood Estimates				Extreme Value Imputation Estimates			
Coefficient	No Delirium	Delirium	Difference	SE	No Delirium	Delirium	Difference	SE
Intercept	58.2	55.8	-2.4	0.6	58.2	55.7	-2.5	0.6
Acute period	-0.5	-1.7	-1.2	0.3	-0.4	-1.5	-1.1	0.4
Post-acute period	1.1	2.2	1.1	0.3	1.1	2.2	1.1	0.3
Intermediate period	1.9	2.0	0.1	0.5	1.0	0.9	0.0	0.4
Long-term cognitive decline	-1.0	-1.4	-0.4	0.1	-0.5	-0.8	-0.3	0.1

Note: The table entries are estimated from the delirium as a covariable model (with respect to the effect of baseline GCP in delirium risk), to more easily evaluate the baseline differences in GCP score according to delirium group.

Study 3: Maximum likelihood vs Complete Case Analysis

Procedure: The third missing data sensitivity analysis only retained participants with complete data at each follow-up time point (N = 241). This is a form of complete case analysis, which is no longer a recommended technique for statistical analysis.[5] Imputation techniques, such as maximum likelihood estimate or multiple imputation, should be used instead.[5] e**Table 1** contains a description of the flow of persons, including deaths and drop-out, at each study time point.

eTable 8. Parameter estimates from components of change model using maximum likelihood estimation and complete cases

	Maximum Likelihood Estimates				Complete Case Estimates			
Coefficient	**No Delirium**	**Delirium**	**Difference**	**SE**	**No Delirium**	**Delirium**	**Difference**	**SE**
Intercept	58.2	55.8	-2.4	0.6	60.0	58.2	-1.8	1.0
Acute period	-0.5	-1.7	-1.2	0.3	-0.3	-1.2	-0.9	0.5
Post-acute period	1.1	2.2	1.1	0.3	1.4	2.2	0.8	0.5
Intermediate period	1.9	2.0	0.1	0.5	1.0	1.6	0.6	0.7
Long-term cognitive decline	-1.0	-1.4	-0.4	0.1	-0.7	-0.9	-0.2	0.2

Note: The table entries are estimated from a the delirium as a covariable model (with respect to the effect of baseline GCP in delirium risk), to more easily evaluate the baseline differences in GCP score according to delirium group.

Results: These results are shown in **Table e8**, where the slope for the no delirium group was less steep than in the maximum likelihood imputation (-0.7 vs -1.0). The effect of delirium was also lower (-0.2 vs -0.4). However, the same general pattern of results held as shown in the maximum likelihood imputation approach.

Summary: Based on these three sensitivity analyses, our results are robust to the way missing data were handled. When using maximum likelihood imputation or multiple imputation, two modern and recommended imputation approaches, our results are nearly identical. When using extreme value imputation and complete case analysis, our estimates for cognitive decline and effect of delirium on cognitive decline lessened, but still held the same pattern of decline and delirium increasing said decline. Based on these findings, we can be confident that missing data does not impact the findings of this study.

eAppendix 4 - Comparing change in non-delirium group to that of the non-surgical comparison group

Objective: Our third sensitivity analysis explored the rate of cognitive decline in the non-surgical comparison group. Because we saw increased rates of decline in this analysis compared to previous findings in the SAGES sample,[1] we wanted to test if similar rates of decline were seen in non-delirium surgical group and the non-surgical comparison group. We did not include the delirium surgical group in this analysis because we knew there was a faster rate of cognitive decline in the delirium group based on the results reported in the main manuscript.

Procedure: For this analysis, we did not use the retest-adjusted GCP score, since the retest adjustment was made based on the non-surgical comparison group. Instead, we used the raw GCP scores. We also fit an adjusted form of the components of change model, based on the available data for the non-surgical comparison group. The adjusted design matrix for this analysis was:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & t_1 & 1 & 0 & 0 \\ 1 & t_2 & 1 & 1 & 0 \\ 1 & t_6 & 1 & 1 & * \\ 1 & t_{12} & 1 & 1 & * \\ 1 & t_{18} & 1 & 1 & 1 \\ 1 & t_{24} & 1 & 1 & 1 \\ 1 & t_{30} & 1 & 1 & 1 \\ 1 & t_{36} & 1 & 1 & 1 \end{bmatrix}$$

The adjusted design matrix still contained five latent variables: intercept, longer-term cognitive decline (slope), acute period and retest, post-acute period, and intermediate period. This model was estimated as a multiple group latent growth curve model, estimated simultaneously in both the surgical and non-surgical groups. The intermediate period only had estimated loadings at months 6 and 12 since data were only available out to 36 months in the non-surgical comparison group. The acute period is referred to as acute period and retest in this model, since only the surgical group experiences acute period . Moreover, post-acute and intermediate periods are only estimated in the surgical group because only the surgical group experiences either of those effects.

Results: The results shown in e**Table 9** suggest the only major difference between the surgical non-delirium and non-surgical comparison group is in the acute period and retest effect, -1.03, SE = 0.27, $p < .001$. However, this result is expected as the difference is showing the acute period effect observed in the surgical group which is not seen in the non-surgical comparison group since they did not have surgery. There was no significant

difference in the longer-term cognitive decline effect between the two groups, 0.03, SE = 0.19, p = 0.89. Based on these results, there is no evidence for a difference in the rate of longer-term cognitive decline in the non-surgical comparison group and surgical non-delirium groups.

eTable 9. Change in cognition from baseline through 36 months in the non-surgical comparison group and surgical non-delirium results.

	Non-surgical comparison group		Surgical non-delirium group	
Coefficient	Estimate	SE	Estimate	SE
Intercept	58.0	0.9	58.5	0.4
Retest & acute period	2.2	0.2	1.1	0.2
Post-acute period	-	-	1.4	0.2
Intermediate period	-	-	1.2	0.3
Long-term period	-0.7	0.1	-0.7	0.1

Notes: SE, standard error. Table entries are parameter estimates from a multiple group growth curve model. Immediate recovery and longer-term recovery effects were estimated in the surgical sample only.

eAppendix 5 – Rates of long-term cognitive decline in SAGES

In the main text, the reported long-term cognitive decline results were in units of GCP per year and standard deviation units per year. These were -1.0 GCP units per year for the non-delirium group (0.14 standard deviation units), and -1.4 GCP units (0.19 standard deviation units) in the delirium group.

The results of 0.14 and 0.19 are calculated by dividing the long-term cognitive decline effect by the SAGES sample standard deviation of the GCP, which is 7.3 (-1.0 / 7.3 = -0.14; -1.4 / 7.3 = -0.19). Using the sample standard deviation to calculate rates of cognitive decline is consistent with previous research on the SAGES sample,[1] which found rates of -0.03 in the non-delirium group and -0.08 in the delirium group when the SAGES cohort was followed out to 36 months.

However, when the GCP was developed, it was scaled to have a population mean of 50 and population standard deviation of 10.[6] Therefore, if the population standard deviation was used to calculate decline in standard deviation units, the results would be declines of 0.10 per year in the non-delirium group and 0.14 per year in the delirium group in the current study, and declines of 0.02 and 0.06 in the previous study following the SAGES sample to 36 months.

Regardless of whether the sample or population standard deviation is used to calculate decline in SD/year, the non-delirium group in the current study (0.10 SD/year using population SD, 0.14 SD/year using sample SD) is close to an estimate of cognitive decline in persons with mild cognitive impairment, which is about 0.10 SD/year.[7] The results for the delirium group (0.14 SD/year using population SD, 0.19 SD/year using sample SD) are above the estimate of MCI and approach an estimate for preclinical Alzheimer's disease, which is 0.21 SD/year.[8]

eReferences

1. Inouye SK, Marcantonio ER, Kosar CM, et al. The short-term and long-term relationship between delirium and cognitive trajectory in older surgical patients. *Alzheimer's & Dementia.* 2016;12(7):766-775.
2. McArdle JJ, Hamagami F. Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: *New methods for the analysis of change.* Washington, DC, US: American Psychological Association; 2001:139-175.
3. Feng Y, Hancock GR, Harring JR. Latent Growth Models with Floors, Ceilings, and Random Knots. *Multivariate Behavioral Research.* 2019;54(5):751-770.
4. Blom G. *Statistical estimates and transformed beta-variables*, Almqvist & Wiksell; 1958.
5. Enders CK. *Applied missing data analysis.* Guilford press; 2010.
6. Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and Validation of an Innovative Approach for Estimating General Cognitive Performance. *Neuroepidemiology.* 2014;42(3):144-153.
7. Johnson JK, Gross AL, Pa J, et al. Longitudinal change in neuropsychological performance using latent growth models: a study of mild cognitive impairment. *Brain Imaging and Behavior.* 2012;6(4):540-550.
8. Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine.* 2000;19(11-12):1555-1566.