Supplementary Appendix

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

Supplement to: Pandharipande PP, Girard TD, Jackson JC, et al. Long-term cognitive impairment after critical illness. N Engl J Med 2013;369:1306-16. DOI: 10.1056/NEJMoa1301372

SUPPLEMENTARY APPENDIX

Long-Term Cognitive Impairment after Critical Illness

Table of Content

BRAIN-ICU (Bringing to light the Risk Factors And Incidence of Neuropsychological dysfunction in ICU survivors) Study Investigators. [++]

PP Pandharipande, TD Girard, JC Jackson, A Morandi, JL Thompson, BT Pun, NE Brummel, CG Hughes, EE Vasilevskis, AK Shintani, KG Moons, SK Geevarghese, A Canonico, LM Boehm, CL Strength, JA Okahashi, AL Kiehl, VA Anderson, MW Baker, J Dunn, RO Hopkins, GR Bernard, RS Dittus and EW Ely

[++]**BRAIN-ICU INVESTIGATORS**

Vanderbilt University School of Medicine, Nashville, TN: PP Pandharipande, TD Girard, JC Jackson, JL Thompson, BT Pun, NE Brummel, CG Hughes, EE Vasilevskis, AK Shintani, SK Geevarghese, LM Boehm, CL Strength, JA Okahashi, AL Kiehl, VA Anderson, GR Bernard, RS Dittus, and EW Ely; *Saint Thomas Hospital, Nashville, TN:* A Canonico, MW Baker, J Dunn; *Hospital Ancelle, Cremona, Italy and Geriatric Research Group, Brescia, Italy:* A Morandi; *Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, Netherlands:* KG Moons; *and Intermountain Medical Center, Murray, UT and Brigham Young University, Provo, UT:* RO Hopkins; *VA Tennessee Valley Geriatric Research and Education Center (GRECC), Nashville, TN:* TD Girard, EE Vasilevskis, RS Dittus, EW Ely, and SM Gordon.

BRAIN-ICU ADVISORY COUNCIL

William R Hazzard: *Puget Sound VA, Seattle, WA*

Edward H Koo: *University of California San Diego, San Diego, CA*

Daniel T Laskowitz: *Duke University Medical Center, Durham, NC*

George M Martin: *University of Washington, Seattle, WA*

1. Methods

A. Definitions of Respiratory Failure and Shock

We enrolled adult patients in a medical or surgical intensive care unit (ICU) receiving treatment respiratory failure or shock (cardiogenic or septic). We considered a patient to be in respiratory failure if they were receiving any of the following treatments at the time of enrollment: invasive mechanical ventilation, noninvasive positive pressure ventilation, continuous positive airway pressure, supplemental oxygen via a nonrebreather mask, or nasal cannula delivering heated high-flow oxygen. We considered a patient to be in cardiogenic shock if they were being treated at the time of enrollment with an intra-aortic balloon pump or any of the following medications administered for acute cardiac dysfunction: dopamine $\geq$ 7.5 mcg/kg/min, dobutamine $\geq$ 5 mcg/kg/min, norepinephrine $\geq$ 5 mcg/min, phenylephrine $\geq$ 75 mcg/min, epinephrine at any dose, milrinone at any dose (if used with another vasopressor), or vasopressin $\geq$ 0.03 units/min (if used with another vasopressor). We considered a patient to be in septic shock when suspected or proven infection was documented in the setting of hypotension being treated with any of the previously listed medications. Patients who were on long-term ventilatory support prior to their acute illness that resulted in the hospitalization qualified for enrollment in this study if they met criteria for shock (as defined above) or they had a new onset

of respiratory failure, defined as either an increase of pressure support of 5 cms H_2O or positive end expiratory pressure of 2 cms H_2O from the patient's baseline ventilatory settings.

B. Assessment of Pre-existing Cognitive Impairment

To assess pre-existing cognitive impairment, we used the Short Form Informant Questionnaire On Cognitive Decline in the Elderly (IQCODE)[1] for patients $\geq$50 years of age and for patients <50 years but with known memory problems. The Short IQCODE is a questionnaire that was developed as a way of identifying the magnitude of cognitive decline from a pre-morbid level using an informant, and it has been repeatedly shown to be effective in identifying the presence of significant cognitive impairment in medical populations and in elderly populations.[2-5] The Short IQCODE consists of a series of 16 questions that are answered by a surrogate with intimate knowledge of the patient, who compares the patient's present cognitive abilities to those 10 years prior. A score of 1 on a question denotes much improvement, a 3 denotes not much change, and a 5 denotes much worse performance. Total score on the 16 questions is then divided by 16 to generate a score ranging from 1 to 5, with higher scores denoting worsening cognitive function. The Short IQCODE has strong psychometric and diagnostic properties,[1] and has been shown to correlate with other informant rating scales[6,7] (e.g., the Blessed Dementia Rating Scale) as well as other cognitive screening tests such as the Mini Mental State Exam (MMSE), the CamCog, and the Abbreviated Mental Screening Test (AMST).[8] It has been validated against neuropathology (sensitivity and specificity of 73% and 75%),[9] Diagnostic and Statistical Manual of Mental Disorders (DSM)-related diagnoses of dementia (sensitivities and specificities across multiple studies in the 80- 90% range)[7,10] neuroimaging studies,[11] and against future decline and mortality.[12-14] The literature has different published thresholds ranging from 3.3 to 3.9 to determine pre-existing cognitive impairment (higher the threshold score, greater the

impairments required to be determined to have cognitive impairment). In our study, patients who scored $\geq$ 3.3 on the Short IQCODE, and therefore had suspected cognitive impairment, were followed up with the Clinical Dementia Rating (CDR; described in next paragraph) Scale. Those with a CDR score >2, suggestive of severe dementia,[15] were excluded, while those with a CDR of $\leq$2, suggestive of mild to moderate pre-existing cognitive impairment, were included in the study. Of the 821 patients enrolled in the study, 92 patients had a Short IQCODE $\geq$3.3 and thus had the CDR administered. Of these, 47 had a score of 0.5 or 1 (minimal impairment), and 45 (5.5% of 821) had scores of 1 or 2 (mild to moderate cognitive impairment).

The CDR is a numeric rating scale used to quantify the severity of dementia symptoms via the use of a structured clinical interview. Individuals are classified across a total of 5 stages (0, 0.5, 1, 2, 3) ranging from "no impairment" to "severe dementia" based on cognitive and functional performance in areas including memory, orientation/judgment and problem solving, community affairs, home and hobbies, and personal care.[16] The CDR is a robust diagnostic tool which has strong inter-rater reliability,[17] concurrent validity (it correlates highly with comprehensive psychometric measures and other widely used rating scales),[18] and predictive validity (it has been shown to be predictive of autopsy confirmed Alzheimer's disease [93%]).[19,20] When used to identify dementia among those with likely cognitive impairment, it has been shown to be nearly 100% sensitive and 100% specific.[21] A recent review[22] identified the CDR as the "best evidenced" of the many dementia rating instruments, and it is the primary dementia rating tool used both in North America and around the world in clinical trials evaluating dementia progression.[23]

The CDR is based on information gained via a semi-structured interview, and as such, it requires clinical and diagnostic skills. Formal training in "CDR Rating" is recommended and

considered the "gold standard" to ensure that the CDR is being administered and scored appropriately. CDR certification is provided by the Washington University Alzheimer's Disease Research Center and involves completion of the Brief Training and Reliability Protocol, which includes an introduction to the CDR by Dr. John Morris, a leading dementia neurologist, videotapes of 3 patient interviews to orient trainees to the CDR, and six videotapes of patient interviews for "reliability certification." Trainees must agree with the expert rater in at least 5 out of 6 evaluations. Training is rigorous and typically takes 2 full days. Passing this training (and receiving formal certification) was a requirement before we allowed our staff to administer the CDR. Though historically we have had very well trained staff, individuals have failed to receive CDR certification on rare occasions and have thus been unable to administer the CDR, highlighting both the difficulty of the training and our commitment to ensuring that this important tool is appropriately used.

Since the CDR was employed in only a select group of patients, we used a Short IQCODE threshold for descriptive purposes to determine if patients had some evidence of pre-existing cognitive impairment. By using a conservative threshold of $\geq$3.6, 6% of our patients may have had pre-existing cognitive impairment. If we used a Short IQCODE $\geq$3.9 threshold, 4% of our patients may have had pre-existing cognitive impairment. Regardless, we used the Short IQCODE as a continuous variable in our multivariable models to adjust for any preexisting cognitive impairment.

C. Using the Confusion Assessment Method for the ICU and the Richmond Agitation-Sedation Scale for Diagnosing Delirium

The Confusion Assessment Method for the ICU (CAM-ICU) consists of four features which assess the following: acute change or fluctuation in mental status (Feature 1), inattention

(Feature 2), disorganized thinking (Feature 3), and an altered level of consciousness (Feature 4).[24] To be diagnosed as delirious, one needs to be a Richmond Agitation-Sedation Scale (RASS)[25] -3 or more awake and be CAM-ICU positive (i.e. with an acute change or fluctuation in mental status [Feature 1], accompanied by inattention [Feature 2] and either disorganized thinking [Feature 3] or an altered level of consciousness [Feature 4]).[24]

D. Calculation of Midazolam and Fentanyl Equivalents

In order to create a single variable for the benzodiazepine and opiate exposure in our study population, we created summary variables based on published potency conversions. Benzodiazepine exposure was expressed in midazolam equivalents, such that 2.5 milligrams (mg) midazolam = 1 mg lorazepam = 5 mg diazepam. Opiate exposure was expressed in fentanyl equivalents, such that 100 micrograms (mcg) fentanyl = 0.75mg hydromorphone = 5mg morphine.[26-28]

E. Definitions of Select Covariates and Rationale

Severe sepsis was defined as sepsis (note: sepsis defined as presence of infection plus at least 2 systemic inflammatory response syndrome features, recorded prospectively and confirmed again by 3 Intensivists [PPP, TDG, EWE] post-ICU stay to assure sepsis had not been an erroneous admission diagnosis) plus any of the following signs of organ dysfunction on a given day: mechanical ventilation, cardiovascular or renal Sequential Organ Failure Assessment (SOFA) score (SOFA) $\geq$ 2, or neurological organ dysfunction, defined as delirium or coma.

Charlson comorbidity index[29] is a score that predicts the ten-year mortality for a patient who may have a range of comorbid conditions. Clinical conditions and associated scores are as follows: 1 point each for: myocardial infarction, congestive heart failure, peripheral vascular disease, dementia, cerebrovascular disease, chronic lung disease, connective tissue disease, ulcer,

chronic liver disease and diabetes; 2 points each for: hemiplegia, moderate or severe kidney disease, diabetes with complication, tumor, leukemia, lymphoma; 3 points for moderate or severe liver disease; and 6 points each for malignant tumor, metastasis, AIDS. Scores are summed to provide a total score to predict mortality. The range of the score is 0-33 (since some categories above are exclusive), with scores of 1-2 associated with approximately a 25% 10-year mortality.

Framingham Stroke Risk Profile[30] is a widely used clinical score based on the prediction of stroke events observed over a 10-year follow-up period in the Framingham Heart Study. Other studies have used this score to predict risk of coronary artery disease as well as cognitive impairment.[30-32] Similar to the Charlson Comorbidity Index, the Framingham Stroke Risk Score accounts for the pre-ICU status that may predispose patients for poor outcomes, including cognitive impairment, and hence both were included as covariates in our model. The Framingham Stroke Risk Score is based on the following risk factors: age, systolic blood pressure, antihypertensive medication, diabetes, cigarette smoking status, history of cardiovascular disease, atrial fibrillation, and left ventricular hypertrophy as determined by an electrocardiogram. The range of the score is 0-30, with most studies showing worsening outcomes in patients with >10 points as compared to those with lower scores.[30-32]

Sequential Organ Failure Assessment (SOFA) score[33] is an organ dysfunction scoring system and is a validated marker of severity of illness over time.[33] The score is based on six different scores, one each for the respiratory, cardiovascular, hepatic, coagulation, renal, and neurological system. For each system, points can range from 0 for no dysfunction to 4 for organ system failure. Thus, totaling the scores for the 6 organ systems provides a SOFA score range from 0 to 24, with higher scores denoting worse organ dysfunction. We used a modified SOFA

score in our regression models, which excluded the neurological components of the SOFA score, since we accounted for coma separately in all our regression models.

Duration of hypoxemia was defined as number of 15-minute epochs (tracked prospectively) during the ICU stay with an oxygen saturation <90% on pulse oximetry.

Rationale for Covariates

All covariates were chosen a priori based on clinical judgment and previous research, due to their expected associations with the outcomes and with ICU delirium and, thus, potential to be confounders. The rationale for covariates chosen is described below:

1. Age is a risk factor for both delirium and long-term cognitive impairment; thus, age could confound the relationship between delirium and long-term cognitive impairment (LTCI).

2. Level of education is a risk factor for both delirium and LTCI; thus, education could confound the relationship between delirium and LTCI.

3. Charlson comorbidity index[29] provides a marker for chronic disease burden and may affect both delirium and LTCI.

4. Pre-existing cognitive impairment according to the Short IQCODE[1] was included in the model as pre-existing cognitive impairment can also affect both delirium and LTCI.

5. Framingham Stroke Risk Profile[30] accounts for cerebrovascular disease risk factors. Given that both delirium and LTCI may occur secondary to cerebral perfusion abnormalities, we adjusted for a marker of stroke risk.

6. Apolipoprotein E (APOE) genotype has been associated with both delirium and Alzheimer's disease, thus a potential confounder.

7. The SOFA score[33] is an indicator of daily severity of illness assessed by number and severity of organ dysfunctions. While the APACHE score provides a single time point severity of illness

at ICU admission, the SOFA score describes a burden of ongoing illness and is therefore a better marker when assessing distant outcomes such as LTCI. We specifically did not use the APACHE, since the SOFA provided similar information at time of admission and thereafter, and we also had accounted for burden of comorbid illness and stroke risk via the Charlson comorbidity score and the Framingham Stroke Risk Profile.

7. Duration of severe sepsis, hypoxemia, and coma- are all potential risk factors for delirium and LTCI; thus to study the independent association of delirium with LTCI, we needed to account for confounding by these variables.

F. Details of Neuropsychological Tests

At 3 and 12 months (± 1month) after hospital discharge, a neuropsychologist or trained psychology professionals assessed patients' global cognition using the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)[34] and patients' executive function using the Trail Making Test Part B (i.e., Trails B).[35]

a. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [34]

The RBANS is a thorough and robust individually administered test designed to assess comprehensively multiple domains of neuropsychological status of adults. It is typically administered in 45 minutes and is composed of 12 subtests (List Learning, Story Memory, Figure Copy, Line Orientation, Digit Span, Coding, Picture Naming, Semantic Fluency, List Recall, List Recognition, Story Recall, and Figure Recall), generating scores in the five cognitive domains (Immediate and Delayed Memory, Attention, Language, and Visuospatial/Constructional Abilities) which combine to form a Global Index Score.

The RBANS has validated population norms (for ages 20-89) for global cognition and domain scores (mean scores are 100 $\pm$ 15) and has been further validated in diverse populations

including those with mild cognitive impairment (MCI), moderate to severe traumatic brain injuries (TBI), vascular dementias, and Alzheimer's Disease (AD), allowing for comparisons of impairment among these conditions.[34,36-39]

Descriptions of Individual RBANS Subtests

i. Immediate Memory (List Learning and Story Memory)

List Learning: This task involves 10 unrelated words which are read to a participant who, in turn, recites these words from memory after each of 4 trials. Story Memory: This task involves a 12-item story which is read to a participant over 2 trials. Participants are asked to provide verbatim recall of the specific details and components of this story.

ii. Visuospatial/Constructional Ability (Figure Copy and Line Orientation)

Figure Copy: This task involves a complex geometric design/figure with 10 distinct components which is shown to participants who are asked to directly copy the figure on a piece of paper. Line Orientation: This task involves a spectrum of 13 lines in specific configurations. For each of 10 items in this test, participants are shown 2 lines and must identify which lines they match from within the set of 13 lines in the spectrum.

iii. Language (Picture Naming and Semantic Fluency)

Picture Naming: This task involves showing participants 10 line drawings of common objects which must be correctly named by participants. Semantic Fluency: This task involves giving participants 60 seconds to generate as many examples as they can from a given semantic category (e.g., fruits and vegetables).

iv. Attention (Digit Span and Coding)

Digit Span: This task involves presenting participants with stimulus items (digits) increasing in length from 2 digits to 9 digits and requiring them to recite them immediately after hearing them,

in the correct order. Coding: This task involves participants filling in digits corresponding to shapes as quickly as they can, using a code key.

v. Delayed Memory (~ 30 Minute Delay) (List Recall, List Recognition, Story Recall, and Figure Recall)

Delayed memory involves: 1) List Learning Recall: free recall of the words from the initial List Learning subtest; 2) List Learning Recognition: yes/no recognition of the words from the initial List Learning subtest; 3) Story Memory Recall: free recall of the story from the initial Story Memory subtest; 4) Figure Recall: free recall of the figure from the initial Figure Copy subtest.

In the tables and figures provided in the supplement, we provide data for the RBANS Global Cognition Score and 3 Index Scores - Immediate Memory, Delayed Memory, and Attention, as these are the primary and secondary outcomes from the RBANS that we selected a priori.

b. Trail Making Test Part B (Trails B) [35]

The Trails B test is a test of executive function (cognitive flexibility and set shifting), requiring an individual to rapidly draw a line between a series of circled numbers and letters in alternating fashion (e.g. 1, A, 2, B, 3, C, etc.) and as quickly as possible.[35] Trails B raw test scores are converted to standardized (age, sex and education-adjusted) T scores, with a mean of 50 and SD of 10.[35]

G. Sample Size and Model Fit

Our independent variables for the linear regression model were delirium duration in the hospital, duration of coma and mean daily doses of benzodiazepines, opiates, propofol, and dexmedetomidine (all continuous variables) plus the following covariates: age, education, Charlson comorbidity index,[29] Short IQCODE,[1] Framingham Stroke Risk Score,[30] mean SOFA score,[33] mean daily doses of haloperidol, days of severe sepsis, and duration of hypoxemia (all continuous), and Apolipoprotein E (APOE) genotype (dichotomous). All continuous variables

were allowed to have nonlinear associations with the outcome with the exceptions of dexmedetomidine and haloperidol, due to low prevalence of use of these medications. We also allowed an interaction between duration of delirium and duration of coma, but if this interaction was clearly not meaningful ($p > 0.20$), we removed it from the model for parsimony. With 448 patients evaluated for the outcome at 3 months and 382 patients evaluated at 12 months, we could fit a model reliably with 44 and 38 degrees of freedom, respectively, based on the general rule that a model must fit no more than m/10 parameters (where m is the effective sample size) to allow for proper multivariable analysis and to be generalizable to future patients.

H. Table S1. Follow-up Rates of Completed Neuropsychological Tests [a]

	3-month follow-up		12-month follow-up	
Neurocognitive test	Patients with Complete Outcomes N	% of Main Cohort (N=448)	Patients with Complete Outcomes N	% of Main Cohort (N = 382)
RBANS Global Cognition Score	374	83%	325	85%
RBANS Immediate Memory Score	410	92%	347	91%
RBANS Delayed Memory	402	90%	342	90%
RBANS Attention Score	386	86%	328	86%
Trails B	386	86%	330	86%

[a] Follow-up assessments were conducted at 3 and 12 months (+/- 1 month) after discharge by a neuropsychologist or trained psychology professionals using a comprehensive cognitive test battery — the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)[34] — and the Trail Making Test Part B (Trails B).[35] Between 8-17% of the cognitive tests were not fully completed by patients in our main cohort: those evaluated at 3 months (N=448) and 12

months (N=382). Multiple imputation was used to impute missing outcomes data at time of regression modeling, since missing data rarely occurs randomly and excluding patients with partial outcomes data would lead to bias.

I. Table S2. Demographic and In-hospital Characteristics of Patients with and without Complete Neuropsychological Outcomes[a]

	Patients with Complete Cognitive Outcomes[b]	Patients with Partial Cognitive Outcomes[b]	P value
	(N=318)	(N=149)	
Age	59 (48-67)	60 (50-72)	0.12
Race, N (%)			0.27
White	284 (89%)	129 (87%)	
African-American	34 (11%)	19 (13%)	
Sex, N (%)			0.012
Male	172 (54%)	62 (42%)	
Female	146 (46%)	87 (58%)	
Type of Insurance, N (%)			0.32
None	22 (7%)	10 (7%)	
Medicaid	19 (6%)	4 (3%)	
Medicare	36 (11%)	20 (13%)	
Medicare + Medicaid	35 (11%)	22 (15%)	
Private	120 (38%)	46 (31%)	
Medicare + Private	86 (27%)	47 (32%)	
Education (years)	12 (12-14)	12 (11-14)	0.05
CSHA Frailty Index, N (%)			<0.001
Very fit	14 (4%)	8 (5%)	
Well	60 (19%)	22 (15%)	
Well; treated comorbid ds	109 (34%)	44 (30%)	
Apparently vulnerable	70 (22%)	23 (15%)	
Mildly frail	40 (13%)	19 (13%)	
Severely frail	1 (0%)	8 (5%)	
FAQ score at enrollment	0 (0, 2)	0 (0, 4)	<0.001
Charlson comorbidity score	2 (1-3)	2 (1-4)	0.41
Framingham stroke risk score	8 (5-14)	10 (6-15)	0.03

Abbreviations: CSHA Frailty Index, Canadian Study of Health and Aging Frailty Scale;[40] FAQ, Functional Activities Questionnaire[41]

[a] Median (interquartile ranges) unless specified

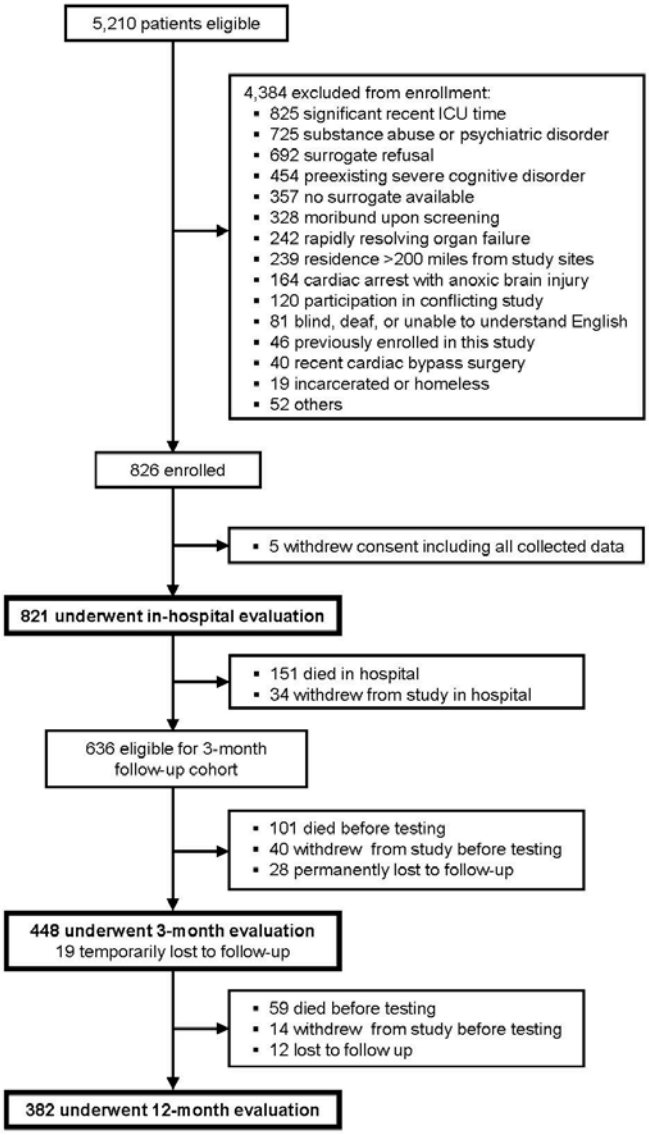
[b] Baseline and in-hospital characteristics for patients with complete outcomes data and for those who had some outcomes data incomplete (i.e. some cognitive tests were missing) at either the 3- or 12-month assessments. Of the 821 patients who had in-hospital data and assessments, 448

patients were assessed at 3 months, and 382 patients were assessed at 12 months; 19 patients were not assessed at 3 months but were available at 12 months, making the total number of unique patients who had follow-up at either one or both time points to be 467.

Though similar in many aspects, some differences did exist (e.g., in education levels, sex distribution, frailty, and activities) between these two groups of patients; thus, the missing outcomes did not occur at random in our cohort, and simply excluding patients with partial data would have led to a selection bias.[42] Multiple imputation strategies were therefore applied at the time of the regression modeling to account for missing data and reduce bias.

2. Results

A. Figure S1. Enrollment and Follow-up



Follow-up cognitive assessments were obtained in 448 patients at 3 months and 382 patients at 12 months after discharge. Nineteen patients who were not reached for testing at 3-month follow-up were tested at 12-month follow-up. Cognitive outcomes were obtained for a total of 467 patients.

B. Table S3. Mortality, Global Cognition and Executive Function according to Age Categories

	≤49 Years N = 192	50-64 Years N = 297	≥65 Years N = 332
Mortality Outcomes, N (%)			
Died in hospital	27 (14%)	52 (18%)	72 (22%)
Died before 3-month follow-up	15 (8%)	33 (11%)	53 (16%)
Died before 12-month follow-up	13 (7%)	21 (7%)	25 (8%)
Cognition Scores at 3 months follow-up, medians (IQR)			
RBANS global cognition score	78 (69-87)	81 (75-87)	77 (67-82)
RBANS immediate memory score	85 (76-100)	85 (76-102)	81 (69-90)
RBANS delayed memory score	83 (74-89)	91 (78-94)	82 (64-93)
RBANS attention score	85 (72-94)	91 (79-100)	85 (72-97)
Cognition Scores at 12 months follow-up, medians (IQR)			
RBANS global cognition score	80 (72-87)	82 (72-89)	80 (70-85)
RBANS immediate memory score	87 (81-103)	87 (76-103)	87 (76-100)
RBANS delayed memory score	83 (71-92)	91 (76-95)	84 (64-95)
RBANS attention score	88 (72-103)	91 (79-100)	88 (78-97)
Executive Function at 3 months follow-up, medians (IQR)			
Trails B score	39 (31-51)	42 (36-48)	39 (32-47)
Executive Function at 12 months follow-up, medians (IQR)			
Trails B score	42 (34-53)	43 (36-50)	42 (35-49)

Long-term cognitive impairment as measured by the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [34] and the Trails B[35] scores occurred similarly in patients ≤49 years, those 50-64 years, and those ≥65 years of age. The lower mortality in patients ≤49 years of age may indicate that the younger patients survived their critical illness but at the expense of having significant cognitive impairment, versus the older patients in whom mortality was the highest.

C. Table S4. Effect of Duration of Delirium and Sedative Exposure on Global Cognition and Executive Function (Complete Case Analysis) [a,b]

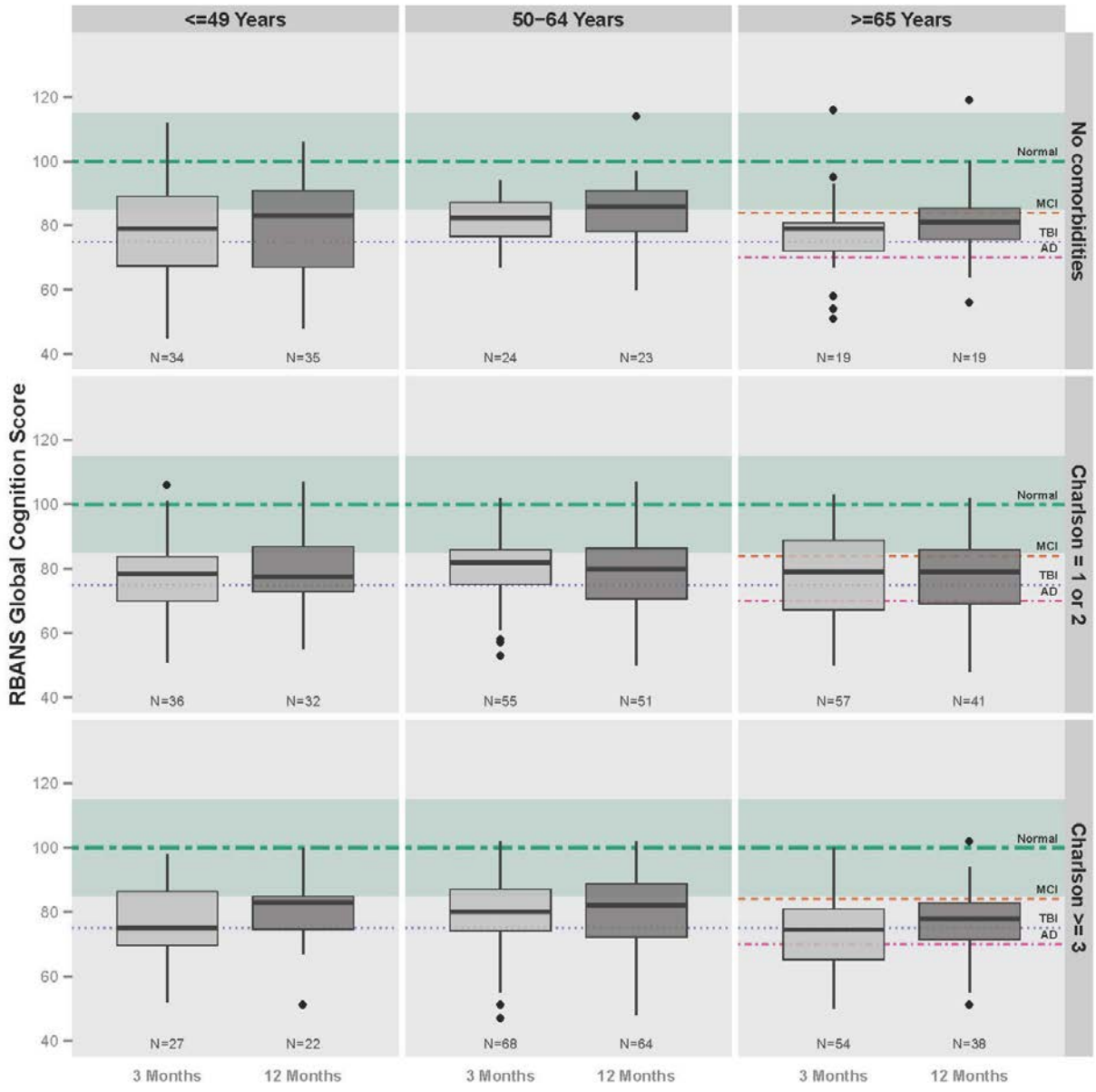
Independent Variables			RBANS Global Cognition Score			
	25th percentile	75th percentile	3-month assessment		12-month assessment	
			Difference in score (95% CI)	*P value*	*Difference in score (95% CI)*	*P value*
Delirium duration (days)	0	5	-7.0 (-11.7, -2.5)	0.001	-5.9 (-10.8, -1.1)	0.10
Coma duration (days)	0	4	1.6 (-5.4, 8.5)	0.29	2.8 (-2.7, 8.2)	0.40
Benzodiazepines (mean daily dose, mg)[c]	0	7.88	0.3 (-3.7, 4.3)	0.20	-0.1 (-4.6, 4.3)	0.41
Propofol (mean daily dose, mg)	0	804	-1.0 (-4.3, 2.2)	0.78	-1.08 (-5.0, 2.9)	0.62
Dexmedetomidine (mean daily dose, mcg)	0	3826	-4.3(-12.8, 4.2)	0.31	-5.7 (-15.4, 3.9)	0.25
Opiates (mean daily dose, mg)[c]	13.3	1238	5.5 (1.3, 9.7)	0.05	1.3 (-3.4, 6.0)	0.33
Independent Variables			**Trails B Executive Function Score**			
			3-month assessment		**12-month assessment**	
			Difference in score (95% CI)	*P value*	*Difference in score (95% CI)*	*P value*
Delirium duration (days)	0	5	-7.3 (-11.9, -2.7)	0.001	-7.5 (-12.5, -2.5)	0.001
Coma duration (days)	0	4	-2.7 (-7.7, 2.3)	0.23	-0.1 (-5.6, 5.4)	0.22
Benzodiazepines (mean daily dose, mg)[c]	0	7.88	-3.4 (-7.7, 0.9)	0.05	0.1 (-4.3, 4.6)	0.08
Propofol (mean daily dose, mg)	0	804	-3.2 (-6.8, 0.4)	0.22	-3.0 (-7.0, 0.9)	0.28
Dexmedetomidine (mean daily dose, mcg)	0	3826	-2.6 (-11.8, 6.6)	0.59	-2.3 (-12.1, 7.5)	0.64
Opiates (mean daily dose, mg)[c]	13.3	1238	7.6 (3.1, 12.2)	0.01	4.5 (-0.2, 9.2)	0.08

[a] Complete case analysis excluded patients that had missing outcomes data (see Table S1). Our main analysis used multiple imputation to account for missing data in 8-17% of cognitive tests, given that missing data is rarely random and excluding these patients introduces selection bias.[42] Nevertheless, sensitivity analyses using only patients with complete outcome data revealed results similar to those in the main cohort.

[b] Results shown are from linear regression models which assess the relationship of delirium, coma and sedative/analgesic medications with RBANS global cognition[34] and Trails B scores,[35] adjusting for a priori chosen confounders: age, education level, Charlson comorbidity index,[29] pre-existing cognitive impairment with the Short IQCODE,[1] the ApoE genotype, the Framingham Stroke Risk Score,[30] and ICU variables including the mean SOFA score,[33] mean haloperidol dose, duration of severe sepsis, duration of hypoxemia, and an interaction of delirium*coma. Difference in scores (point estimates) in the RBANS and the Trails B scores in the linear regression analysis reflect a comparison between the 25th and the 75th percentile values for each variable (with the exception of dexmedetomidine dose; because over 85% of patients received no dexmedetomidine, we used the minimum and maximum doses). For example, when comparing patients with 0 and 5 days of delirium (the 25th and 75th percentile values of delirium duration in our cohort), holding all other covariates constant at their median or modes, those with 5 days of delirium had, on average, RBANS global scores 7 points lower at 3 months and 5.9 points lower at 12 months than patients with no delirium. Each of these is reflective of approximately a 0.5 SD decrease in scores, which is clinically significant.[43,44]

[c] Exposure to sedative and analgesic medications was defined as the mean daily dose of benzodiazepines, propofol, dexmedetomidine, and opiates during the ICU stay. Benzodiazepine exposure is expressed in midazolam equivalents, such that 2.5 mg midazolam = 1 mg lorazepam = 5 mg diazepam. Opiate exposure is expressed in fentanyl equivalents, such that 100 micrograms (mcg) fentanyl = 0.75 mg hydromorphone = 5 mg morphine.[26-28]

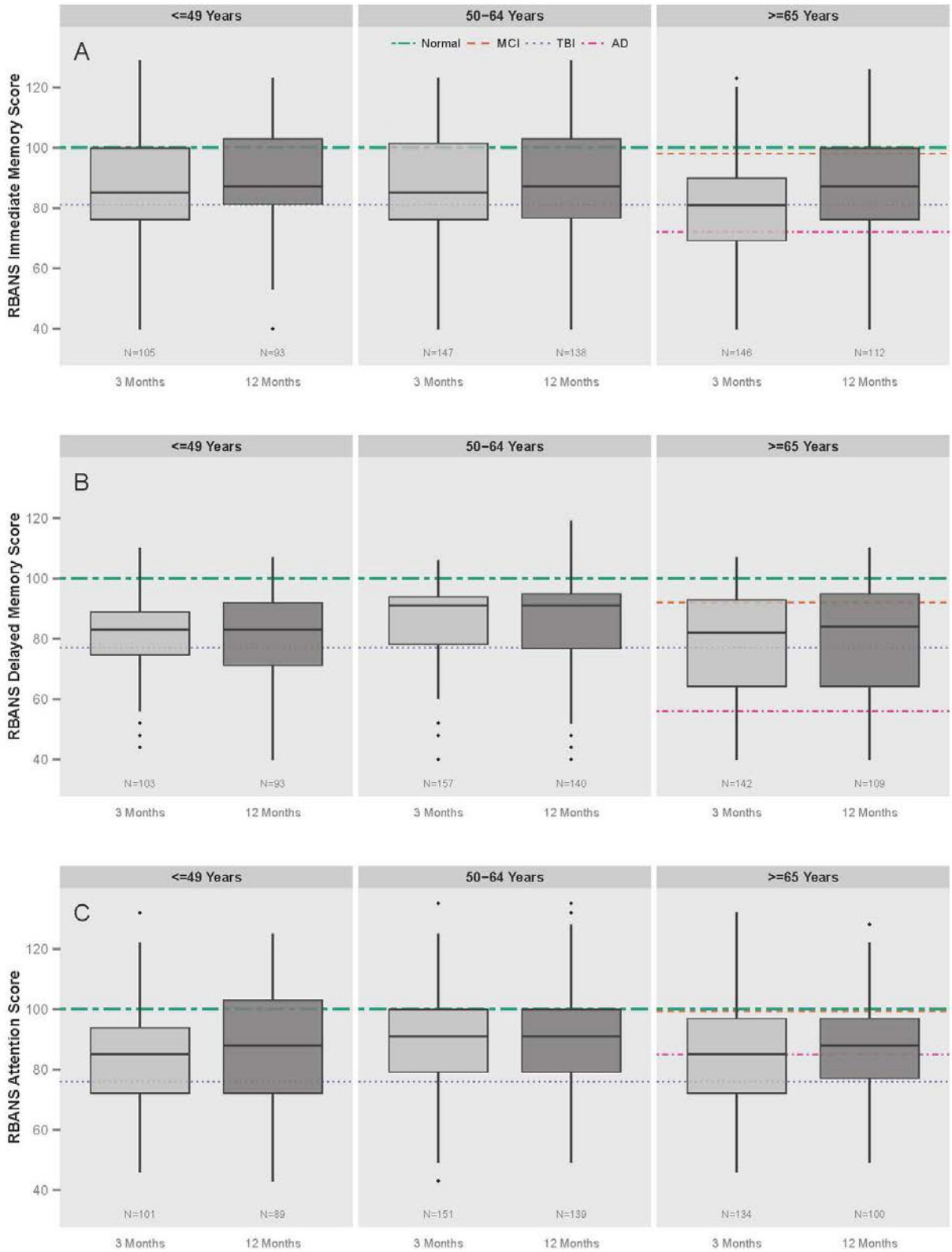
D. Figure S2. Global Cognition Scores by Age and Baseline Comorbidity



These box-and-whisker plots display the RBANS global cognition scores in our cohort at 3-month (light gray boxes) and 12-month follow-up (dark gray boxes) according to age and number of comorbidities: outcomes for patients ≤49 years in the left panels, for those 50-64 years in the middle panels, and for those ≥65 years in the right panels; outcomes for patients with no comorbidities (Charlson comorbidity index score of 0) in the top panels, for patients with

some comorbidities (Charlson comorbidity index score of 1 or 2) in the middle panels, and for patients with significant comorbidities (Charlson comorbidity index score of ≥3) in the lower panels. For each box and whisker plot, horizontal bars indicate the median, upper and lower limits of the boxes indicate interquartile range (IQR), and ends of whiskers indicate 1.5*IQR. Data points falling outside the whiskers are shown as black dots. Green lines in each panel show the age-adjusted population mean (100) for healthy adults, and the green band reflects the age-adjusted standard deviation (15). Other lines indicate population means for those with mild cognitive impairment (MCI), moderate traumatic brain injury (TBI), or with mild Alzheimer's disease (AD). Population means for MCI and AD are only shown for the age ≥65 category, since RBANS population norms for these entities have only been generated in that age group. The number of patients with complete data at each time point is shown under each plot; these counts do not include patients who underwent partial outcomes assessment and thus required imputation of outcomes during the analyses (see methods). **Interpretation**: Older and younger patients, with or without comorbid diseases, had deficits in the global cognition scores at follow-up. Even in the category of patients 49 years and younger with a Charlson Comorbidity score of 0 at baseline, for example, more than a third had cognitive function at the 12-month follow-up testing that was commensurate with that of moderate TBI patients, and a quarter had results similar to patients with mild Alzheimer's disease.

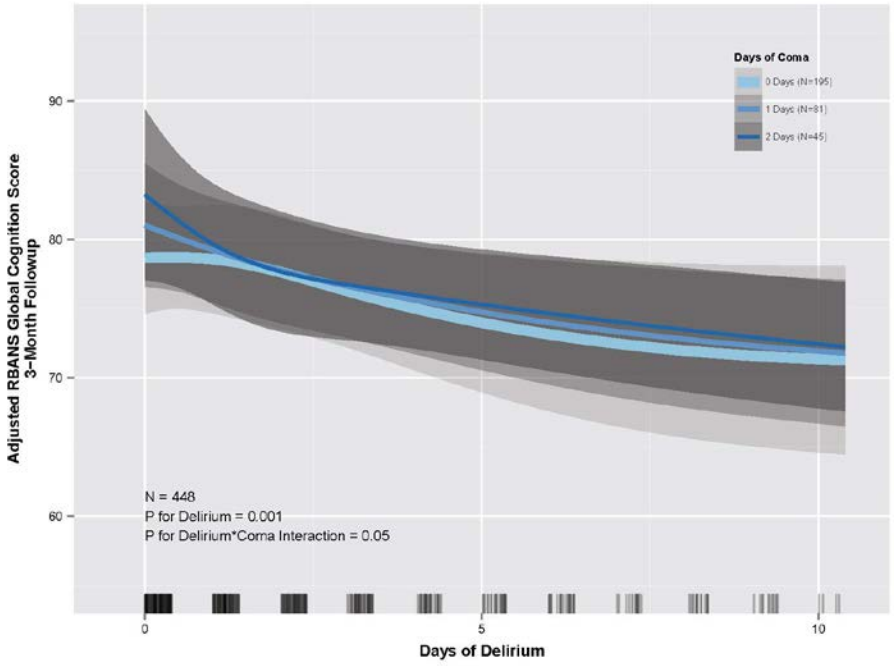
E. Figure S3. Individual RBANS Domain Scores in Survivors of Critical Illness



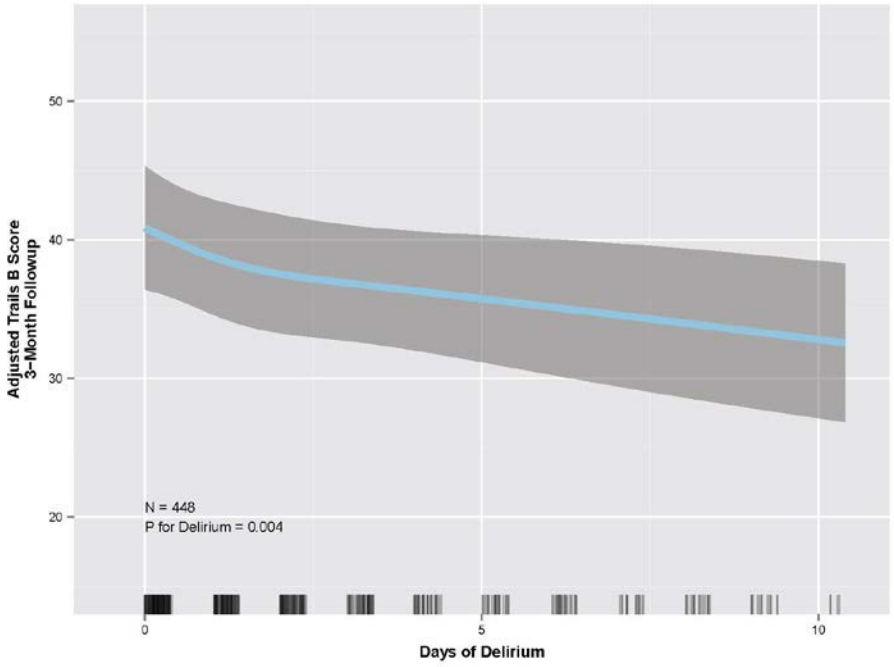
These box-and-whisker plots show the RBANS individual domain scores for immediate memory (Panel A), delayed memory (Panel B), and attention (Panel C) in our follow-up cohort at 3-month (light gray boxes) and 12-month follow-up (dark gray boxes) according to age: outcomes for patients ≤49 years on the left panels, for those 50-64 years in the middle, and for those ≥65 years on the right panels. For each box and whisker plot, horizontal bars indicate the median, upper and lower limits of the boxes indicate interquartile range (IQR), and ends of whiskers indicate 1.5*IQR. Data points falling outside the whiskers are shown as black dots. Green lines in each panel show the age-adjusted population mean (100) for healthy adults. Other lines indicate expected population means for those with mild cognitive impairment (MCI), moderate traumatic brain injury (TBI), or mild Alzheimer's disease (AD) based on other cohort studies. Expected population means for MCI and AD are only shown for the age ≥65 category, since RBANS population norms for these conditions have only been generated in that age group. The number of patients with complete data at each time point is shown under each plot; these counts do not include patients who underwent partial outcomes assessment and thus required imputation of outcomes during the analyses (see methods). **Interpretative example**: Panel A shows RBANS immediate memory scores in our tested cohort separately for those ≤49 years (on left), 50-64 (middle) and for those ≥65 years (on right), at 3 and 12 months. In patients ≥65 years of age, for example, horizontal bars and upper and lower limits of the box show the median (IQR) of RBANS immediate memory scores corresponding to 81 (69, 90) and 87 (76, 100), at 3 and 12 months, respectively. These median scores are similar to those seen in moderate TBI as seen by the dashed line representing the population mean for TBI around a score of 80.

F. Figure S4. Duration of Delirium and Global Cognition Scores (Panel A) and Executive Function Scores (Panel B) at 3-Month Follow-Up

Panel A

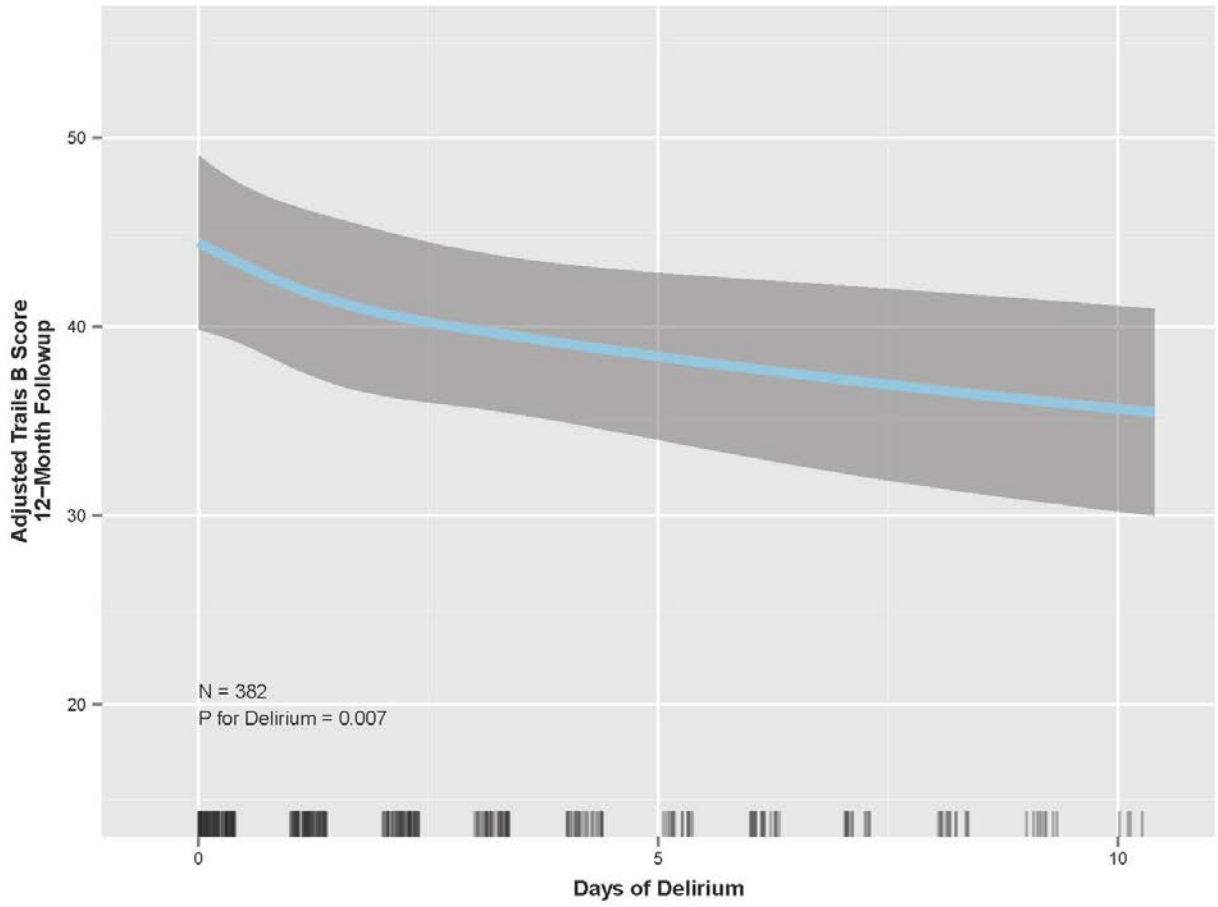


Panel B



Longer duration of delirium (X-axis) was independently associated with worse RBANS global cognition scores (Panel A) and Trails B executive function scores (Panel B) at 3-month follow-up. RBANS global cognition scores have age-adjusted population means of 100 (±15). Trails B scores are T scores with an age, sex, and education-adjusted mean of 50 (±10). Rug plots along the X-axis show the distribution of patients according to duration of delirium. Although delirium could be assessed for up to 30 days in the study, the X-axis is truncated at 10 days, since 90% of our patients had a duration of delirium ≤10 days; all available data were used in the multivariable modeling. **Panel A**: The relationship between delirium and global cognition scores at 3 months was modified by the duration of coma (P for interaction=0.05). The three colored lines in the figure reflect the relationship between delirium and RBANS global cognition scores for patients with 0 days, 1 day, and 2 days of coma, respectively. Line thickness represents the number of patients with the given duration of coma, with thicker lines indicating more patients. Patients with a longer duration of coma had a greater decrement in RBANS global cognition score for each day of delirium; while statistically significant, this interaction may not be clinically significant. **Panel B:** Longer duration of delirium was independently associated with lower Trails B executive function scores at 3-months follow-up. **Interpretative example**: when comparing patients with 0 and 5 days of delirium (the 25th and 75th percentile values of delirium duration in our cohort), holding all other covariates constant at their median or mode, those with 5 days of delirium had, on average, 5.1 points lower Trails B scores at 3-month follow-up than those with 0 days of delirium, a 0.5 SD change that is clinically significant.

G. Figure S5: Duration of Delirium and Executive Function Scores at 12-Month Follow-Up



Longer durations of delirium (X-axis) were independently associated with worse Trails B executive function scores at 12-month follow-up. Trails B scores are T-scores with an age, sex, and education-adjusted mean of 50 (±10). Rug plots along the X-axis show the distribution of delirium durations. Although delirium could be assessed for up to 30 days in the study, the X-axis is truncated at 10 days, since 90% of patients had duration of delirium ≤10 days; all available data were utilized in the multivariable modeling.

H. Overlap between Impairments at 3 and 12 months

In order to understand the temporal course and overlap of cognitive impairment in our patients, the RBANS global scores were dichotomized with scores <78 reflective of overt cognitive impairment (1.5 SD below normal). Of the 374 patients with complete RBANS global scores, 167 (45%) were impaired at 3 months. Of these patients, 26% were not tested at 12 months due to death, withdrawal, or incomplete outcomes data; 22% were no longer impaired; and 51% were still impaired. 8% of patients who were not impaired at 3 months were classified as cognitively impaired at 12 months.

I. Duration of Delirium and RBANS domain scores.

Longer delirium duration was a risk factor for worse function in several individual RBANS domains (population mean scores for the RBANS subscales is also 100 + 15). An increase from 0 to 5 days of delirium was associated with lower scores in attention at 3 months (-8.3 [95% confidence interval -14.4 to -2.1], P<0.001) and 12 months (-11.3 [-17.1 to -5.5], P=0.002), in immediate memory at 3 months (-6.8 [-12.8 to -0.8], P=0.05), and delayed memory at 3 months (-5.9 [-10.6 to -1.3], P=0.01). Each of these is reflective of approximately a 0.5-1SD decrease in scores, which is clinically and statistically significant.[43,44]

J. Delirium, Frailty, Mechanical Ventilation, and Cognitive Impairment

In our linear regression model, we adjusted for a number of confounders which relate to patient illness and frailty. These included age, the Charlson comorbidity index, the Short IQCODE for pre-existing cognitive impairment, the Framingham Stroke Risk Score, the mean SOFA score, and the days of severe sepsis. Delirium was the most strongly associated risk factor of long-term cognitive impairment. None of the other covariates listed above were consistent in their association with cognitive impairment. It is likely that general "frailty" is indeed on the causal

pathway between delirium and worsening cognition (or is a consequence of worsening cognition). We therefore did not adjust for frailty in our multivariable analysis because it may have falsely masked the effect of delirium. We did look at correlations between frailty and cognitive outcomes and found a moderate correlation between higher (worse) frailty scores and lower (worse) RBANS global scores at 3- and 12-month follow-up. With Spearman's rho ranging from -0.26 to -0.32, this suggests that the relationship is meaningful, though certainly not a major predictor of long-term cognitive impairment.

In our prior pilot work by Girard et al,[45] which showed an association between delirium and LTCI, we analyzed our data in 3 separate ways to determine the role of mechanical ventilation in LTCI. First, we studied the role of delirium and LTCI, adjusting for covariates, and found that duration of delirium was associated with LTCI. Next, we substituted duration of delirium with duration of mechanical ventilation and found that duration of mechanical ventilation was not associated with LTCI. Finally, we added both duration of mechanical ventilation and delirium into the model and found that duration of delirium continued to be associated with LTCI and mechanical ventilation was not.

K. Table S5. Additional Demographic Data

	In-hospital cohort (N=821)	**Follow-up cohort (N=467)**
Variable		
Pre-existing cognitive impairment measured by Short IQCODE	3 (3-3)	3 (3-3)
Framingham stroke risk[a]	9 (6-14)	9 (5-14)
Delirium days (among those with delirium)	4 (2-7)	3 (2-7)
Coma days (among those with coma)	3 (2-6)	3 (1-5)
Severe sepsis during study period, N (%)	572 (70%)	297 (64%)
ICU length of stay (days)	5 (3-11)	5 (2-10)
Duration of benzodiazepine therapy (among exposed), days	3 (1-6)	3 (1-6)
Duration of propofol therapy (among exposed), days	2 (1-4)	2 (1-4)
Duration of opiate therapy (among exposed), days	3 (2-7)	3 (2-7)
Duration of dexmedetomidine therapy (among exposed), days	2 (1-3)	2 (1-3)

[a]Framingham stroke risk- score ranges from 0 to 30 points, with scores >10 associated with worse cardiac and neurological outcomes.

L. Study Oversight: This BRAIN-ICU observational study was designed by members of the steering committee (EWE, PPP, TDG, JCJ, AKS, KGM, ROH, GRB and RSD) with guidance from the advisory council (WRH, EHK, DTL, GMM). Data were gathered by qualified research nurses and psychologists, and analyzed by biostatisticians. The steering committee vouches for the data, the analysis, and the decision to submit the manuscript for publication. The first author and last author drafted the initial manuscript with other authors providing edits thereafter.

M. References

1. Jorm AF. The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. Int Psychogeriatr 2004;16:275-93.
2. Louis B, Harwood D, Hope T, Jacoby R. Can an informant questionnaire be used to predict the development of dementia in medical inpatients? Int J Geriatr Psychiatry 1999;14:941-5.
3. Li F, Jia XF, Jia J. The Informant Questionnaire on Cognitive Decline in the Elderly individuals in screening mild cognitive impairment with or without functional impairment. J Geriatr Psychiatry Neurol 2012;25:227-32.
4. Mailles A, De Broucker T, Costanzo P, et al. Long-term outcome of patients presenting with acute infectious encephalitis of various causes in France. Clin Infect Dis 2012;54:1455-64.
5. Waltrowicz W, Ames, D., McKenzie, S. and Flicker, L. Burden and stress on relatives (informal carers) of dementia sufferers in psychogeriatric nursing homes. Australian Journal on Ageing 1996;15:115-8.
6. Pisani MA, Inouye SK, McNicoll L, Redlich CA. Screening for preexisting cognitive impairment in older intensive care unit patients: use of proxy assessment. J Am Geriatr Soc 2003;51:689-93.
7. Morales JM, Bermejo F, Romero M, Del-Ser T. Screening of dementia in community-dwelling elderly through informant report. Int J Geriatr Psychiatry 1997;12:808-16.
8. Stratford JA, LoGiudice D, Flicker L, Cook R, Waltrowicz W, Ames D. A memory clinic at a geriatric hospital: a report on 577 patients assessed with the CAMDEX over 9 years. Aust N Z J Psychiatry 2003;37:319-26.
9. Thomas LD, Gonzales, M. F., Chamberlain, A., Beyreuther, K., Masters, C. L., Flicker, L. . Comparison of clinical state, retrospective informant interview and the neuropathologic diagnosis of Alzheimer's disease. International Journal of Geriatric Psychiatry 1994;9.
10. Lim HJ, Lim JP, Anthony P, Yeo DH, Sahadevan S. Prevalence of cognitive impairment amongst Singapore's elderly Chinese: a community-based study using the ECAQ and the IQCODE. Int J Geriatr Psychiatry 2003;18:142-8.
11. Cordoliani-Mackowiak MA, Henon H, Pruvo JP, Pasquier F, Leys D. Poststroke dementia: influence of hippocampal atrophy. Arch Neurol 2003;60:585-90.
12. Barba R, Martinez-Espinosa S, Rodriguez-Garcia E, Pondal M, Vivancos J, Del ST. Poststroke dementia : clinical features and risk factors. Stroke 2000;31:1494-501.
13. Henon H, Durieu I, Guerouaou D, Lebert F, Pasquier F, Leys D. Poststroke dementia: incidence and relationship to prestroke cognitive decline. Neurology 2001;57:1216-22.
14. Henon H, Lebert F, Durieu I, et al. Confusional state in stroke: relation to preexisting dementia, patient characteristics, and outcome. Stroke 1999;30:773-9.
15. Berg L. Clinical Dementia Rating (CDR). Psychopharmacol Bull 1988;24:637-9.
16. Morris JC, Ernesto C, Schafer K, et al. Clinical dementia rating and reliability in multicenter studies: the Alzheimer's disease cooperative study experience. Neurology 1997;48 1508-10.
17. Burke WJ, Miller JP, Rubin EH, et al. Reliability of the Washington University Clinical Dementia Rating. Arch Neurol 1988;45:31-2.
18. Choi SH, Lee BH, Kim S, et al. Interchanging scores between clinical dementia rating scale and global deterioration scale. Alzheimer Dis Assoc Disord 2003;17:98-105.

19. Berg L, McKeel DW, Jr., Miller JP, et al. Clinicopathologic studies in cognitively healthy aging and Alzheimer's disease: relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. Arch Neurol 1998;55:326-35.
20. Galvin JE, Powlishta KK, Wilkins K, et al. Predictors of preclinical Alzheimer disease and dementia: a clinicopathologic study. Arch Neurol 2005;62:758-65.
21. Chaves ML, Camozzato AL, Godinho C, et al. Validity of the clinical dementia rating scale for the detection and staging of dementia in Brazilian patients. Alzheimer Dis Assoc Disord 2007;21:210-7.
22. Rikkert MG, Tona KD, Janssen L, et al. Validity, reliability, and feasibility of clinical staging scales in dementia: a systematic review. Am J Alzheimers Dis Other Demen 2011;26:357-65.
23. Schafer KA, Tractenberg RE, Sano M, et al. Reliability of monitoring the clinical dementia rating in multicenter clinical trials. Alzheimer Dis Assoc Disord 2004;18:219-22.
24. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). JAMA 2001;286:2703-10.
25. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002;166:1338-44.
26. Gammaitoni AR, Fine P, Alvarez N, McPherson ML, Bergmark S. Clinical application of opioid equianalgesic data. Clin J Pain 2003;19:286-97.
27. Anderson R, Saiers JH, Abram S, Schlicht C. Accuracy in equianalgesic dosing. conversion dilemmas. J Pain Symptom Manage 2001;21:397-406.
28. Dundee JW, McGowan WA, Lilburn JK, McKay AC, Hegarty JE. Comparison of the actions of diazepam and lorazepam. Br J Anaesth 1979;51:439-46.
29. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83.
30. Llewellyn DJ, Lang IA, Xie J, Huppert FA, Melzer D, Langa KM. Framingham Stroke Risk Profile and poor cognitive function: a population-based study. BMC Neurol 2008;8:12.
31. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. Stroke 1994;25:40-3.
32. Kannel WB, Dannenberg AL, Abbott RD. Unrecognized myocardial infarction and hypertension: the Framingham Study. AmHeart J 1985;109:581-5.
33. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286:1754-8.
34. Randolph C, Tierney MC, Mohr E, Chase TN. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. J Clin Exp Neuropsychol 1998;20:310-9.
35. Reitan RM, Wolfson D. The Halstead Reitan Neuropsychological Test Battery. Tuscon, AZ: Neuropsychology Press; 1985.
36. Duff K, Hobson VL, Beglinger LJ, O'Bryant SE. Diagnostic accuracy of the RBANS in mild cognitive impairment: limitations on assessing milder impairments. Arch Clin Neuropsychol 2010;25:429-41.

37. Duff K, Humphreys Clark JD, O'Bryant SE, Mold JW, Schiffer RB, Sutker PB. Utility of the RBANS in detecting cognitive impairment associated with Alzheimer's disease: sensitivity, specificity, and positive and negative predictive powers. Arch Clin Neuropsychol 2008;23:603-12.
38. McKay C, Casey JE, Wertheimer J, Fichtenberg NL. Reliability and validity of the RBANS in a traumatic brain injured sample. ArchClinNeuropsychol 2007;22:91-8.
39. Hobson VL, Hall JR, Humphreys-Clark JD, Schrimsher GW, O'Bryant SE. Identifying functional impairment with scores from the repeatable battery for the assessment of neuropsychological status (RBANS). Int J Geriatr Psychiatry 2010;25:525-30.
40. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005;173:489-95.
41. Pfeffer RI, Kurosaki TT, Harrah CH, Jr., Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323-9.
42. Roderick J, Little A, Rubin DB. Statistical Analysis with Missing Data: Wiley; 1987.
43. Harvey PD. Clinical applications of neuropsychological assessment. Dialogues in clinical neuroscience 2012;14:91-9.
44. Podell Kd, P.A.; Barrett, P.; McCullen, A.; Goldbert, E. The Handbook of Psychology. Hoboken, New Jersey John Wiley and Sons; 2003.
45. Girard TD, Jackson JC, Pandharipande PP, et al. Delirium as a predictor of long-term cognitive impairment in survivors of critical illness. Crit Care Med 2010;38:1513-20.