



Published in final edited form as:

*Neurocrit Care*. 2021 April ; 34(2): 403–412. doi:10.1007/s12028-020-01126-8.

## The Feasibility and Validity of Objective and Patient-Reported Measurements of Cognition during Early Critical Illness Recovery

Matthew B. Maas, MD, MS<sup>1,2</sup>, Bryan D. Lizza, PharmD, MS<sup>3</sup>, Minjee Kim, MD<sup>1,2</sup>, Maged Gendy, BS<sup>1,2</sup>, Eric M. Liotta, MD, MS<sup>1</sup>, Kathryn J. Reid, PhD<sup>1,2</sup>, Phyllis C. Zee, MD, PhD<sup>1,2</sup>, James W. Griffith, PhD<sup>4</sup>

<sup>1</sup>Department of Neurology, Northwestern University, Chicago, IL

<sup>2</sup>Center for Circadian and Sleep Medicine, Northwestern University, Chicago, IL

<sup>3</sup>Department of Pharmacy, Barnes-Jewish Hospital, St. Louis, MO

<sup>4</sup>Department of Medical Social Sciences, Northwestern University, Chicago, IL

### Abstract

**Background:** Cognitive outcomes are an important determinant of quality of life after critical illness, but methods to assess early cognitive impairment and cognition recovery are not established. The objective of this study was to assess the feasibility and validity of objective and patient-reported cognition assessments for generalized use during early recovery from critical illness.

**Methods:** Patients presenting from the community with acute onset of either intracerebral hemorrhage (ICH) or sepsis as representative neurologic and systemic critical illnesses. Early cognitive assessments comprised the Glasgow Coma Scale (GCS), three NIH Toolbox cognition measures (Flanker Inhibitory Control and Attention Test, List Sorting Working Memory Test and Pattern Comparison Processing Speed Test) and two Patient Reported Outcomes Measurement Information System (PROMIS) cognition measures (Cognition-General Concerns and Cognition-Abilities) performed seven days after intensive care unit discharge or at hospital discharge, whichever occurred first.

**Results:** We enrolled 91 patients (53 with sepsis, 38 with ICH), and after attrition principally due to deaths, cognitive assessments were attempted in 73 cases. Median [interquartile range] Sequential Organ Failure Assessment scores for patients with sepsis was 7 [3, 11]. ICH cases included 13 lobar, 21 deep and 4 infratentorial hemorrhages with a median [IQR] ICH Score 2 [1, 2]. Patient-reported outcomes were successfully obtained in 42 (58% overall, 79% of sepsis and

---

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

**Corresponding Author:** Matthew Maas, 625 N Michigan Ave, Suite 1150, Chicago, IL 60611, Phone: 312-503-0479, Fax: 312-503-0056, mbmaas@northwestern.edu.

**Publisher's Disclaimer:** This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

The work was performed at Northwestern University and Northwestern Memorial Hospital.

34% of ICH) patients but scores were anomalously favorable (median 97<sup>th</sup> percentile compared to the general adult population). Analysis of the PROMIS item bank by four blinded, board-certified academic neurointensivists revealed a strong correlation between higher severity of reported symptoms and greater situational relevance of the items ( $\rho = 0.72$ ,  $p=0.002$  correlation with expert item assessment), indicating poor construct validity in this population. NIH Toolbox tests were obtainable in only 9 (12%) patients, all of whom were unimpaired by GCS (score 15) and completed PROMIS assessments. Median scores were 5<sup>th</sup> percentile (interquartile range [2<sup>nd</sup>, 9<sup>th</sup>] percentile) and uncorrelated with self-reported symptoms. Shorter intensive care unit length of stay was associated with successful testing in both patients with ICH and sepsis, along with lower ICH Score in patients with ICH and absence of premorbid dementia in patients with sepsis (all  $p<0.05$ ).

**Conclusions:** Methods of objective and patient-reported cognitive testing that have been validated for use in patients with chronic medical and neurologic illness were infeasible or yielded invalid results among a general sample of patients in this study who were in early recovery from neurologic and systemic critical illness. Longer critical illness duration and worse neurocognitive impairments, whether chronic or acute, reduced testing feasibility.

### Keywords

cognition; critical illness; outcomes

---

### Introduction

Encephalopathy is the primary organ failure syndrome of the central nervous system during critical illness.<sup>1</sup> Acute encephalopathy does not fully resolve in many patients despite resolution of the inciting illness, persisting as static encephalopathy, more commonly called chronic cognitive impairment. Acute care hospitalization, either for critical or non-critical illness, is associated with later cognitive impairments.<sup>2</sup> Delirium is a common manifestation of mild-to-moderate encephalopathy that affects at least half of critically ill patients, and is an independent predictor of worse outcomes including higher mortality and long term cognitive impairments.<sup>3,4</sup> For critical illness survivors, cognitive impairment is a common sequella and important contributor to Post-Intensive Care Syndrome (PICS).<sup>5</sup> Early cognitive testing after acute illness stabilization has been proposed as a method to elucidate risk factors for poor cognitive outcomes and identify opportunities for intervention, currently implemented in at least one large, ongoing critical illness outcome study, but the validity of existing approaches to cognitive testing in this context is not established.<sup>6</sup>

Several validated instruments are available to assess for delirium, but the most reliable tests are binary rather than quantitative, and the utility of delirium screening is uncertain outside the context of acute illness.<sup>1</sup> Acute encephalopathy can be quantified with instruments like the Glasgow Coma Scale (GCS) and Richmond Agitation Sedation Scale.<sup>1</sup> Those scales were developed to measure the severity of impaired consciousness or sedation intensity and are useful as methods for monitoring critically ill patients and for prognostication in composite severity measures, but ceiling effects make them uninformative in the range of impairment seen outside of intensive care units (ICUs).<sup>7-14</sup> The National Institutes of Health supported the development, validation and dissemination of objective (NIH Toolbox) and

patient-reported (Patient Reported Outcomes Measurement Information System; PROMIS) measures of cognition in a format feasible for efficient implementation in clinical research, but those instruments have not yet been validated in post-critical illness patients.<sup>15–17</sup> The objective of this study was to assess the feasibility and validity of objective and patient-reported cognition measurements using NIH Toolbox and PROMIS for generalized use in patients during early recovery from critical illness.

## Methods

### Patients

Patients presenting to ICUs at Northwestern Memorial Hospital between April 2014 and December 2018 were prospectively enrolled in an observational cohort study. The study was conducted with the overall objective of evaluating circadian rhythm changes during critical illness, including effects on sleep function and cognition during early recovery.<sup>18,19</sup> The study was approved by the Institutional Review Board and written informed consent was obtained from patients or their legally authorized representative. We enrolled patients 18 years old with either spontaneous intracerebral hemorrhage (ICH) as a representative neurologic critical illness or acute sepsis as a representative systemic critical illness. We restricted inclusion to patients presenting emergently from the community with demonstrably acute onset of symptoms and excluded patients unlikely to survive for at least 24 hours or unlikely to require at least 48 hours of ICU care, those with baseline need for renal replacement therapy or with hemoglobin concentrations less than 7 g/dL. Enrollment occurred within 24 hours of emergency department presentation. Standard disease-specific and general severity measurements were recorded including GCS, ICH Score and Sequential Organ Failure Assessment (SOFA). Global functional impairment was measured by the modified Rankin Scale (mRS) in which scores range from 0 (asymptomatic) to 6 (dead). The Confusion Assessment Method was used to identify delirium, which was categorized as “confusion” on the GCS when relevant. Premorbid dementia was defined as a history of cognitive impairment based on family interview and medical record review consistent with methods used in calculating the FUNC Score, and baseline disability was measured with the modified Rankin Scale using a validated interview with the patient and family along with medical record review, methods which we have used and reported previously.<sup>11,20–22</sup>

### Measurements

The hospital course of enrolled patients was tracked, and for patients surviving critical illness, a cognitive assessment was performed in the hospital seven days after transferring out of the intensive care unit, or within a day of hospital discharge if discharge occurred earlier. The goal of the early assessment was to establish the patient’s post-critical illness baseline, consistent with methods for the same objective used elsewhere.<sup>6</sup> We obtained patient-reported cognitive assessments using the PROMIS Applied Cognition-General Concerns and Applied Cognition-Abilities instruments. The PROMIS Applied Cognition General Concerns instrument evaluates the extent to which cognitive impairments interfere with functioning, are noticeable to others and impact quality of life, whereas the Abilities instrument focuses on whether cognitive impairments interfere with the ability to perform cognitive tasks, with scores referenced to the U.S. general population.<sup>16</sup> The PROMIS

instruments were obtained by computer adaptive testing (CAT) which is designed to yield good estimate precision with less required responses. In the case of PROMIS CAT, algorithm stopping rules are constrained to a minimum of 4 and maximum of 12 items.<sup>23,24</sup> Objective cognitive testing was performed with NIH Toolbox using the Flanker Inhibitory Control and Attention Test to assess attention and executive functioning, the List Sorting Working Memory Test to assess working memory function, and the Pattern Comparison Processing Speed Test to assess processing speed, as we have previously described.<sup>25</sup> PROMIS and NIH Toolbox test results are reported as T-Scores in which 50 is the reference population mean and 10 is the standard deviation. For NIH Toolbox, we used fully corrected T-Scores that were adjusted for age, gender, race/ethnicity and education. Higher scores for PROMIS Cognition General Concerns indicate greater burden of cognitive impairment symptoms (worse cognitive function). Higher scores for NIH Toolbox tests indicate better cognitive function.

### Item Analyses for Patient-Reported Cognition Questions

The PROMIS cognition instruments were developed for a general population and elicit responses by asking respondents the frequency with which they have experienced the symptom described within the prior week according to the following response scale: very often (several times a day), often (about once a day), sometimes (two or three times), rarely (once), or never. On their face, the construct of some questions appeared likely to be situationally relevant to patients in an acute care environment (e.g. “My thinking has been slow.” or “It has seemed like my brain was not working as well as usual.”), whereas others seemed to have low situational relevance (e.g. “I have made mistakes when writing down phone numbers.” or “I have had trouble remembering where I put things, like my keys or my wallet.”). Expert assessment of questions in patient-reported outcome instrument item banks is helpful in establishing relevance and clinical meaningfulness and was used in the development of PROMIS instruments.<sup>26</sup> We conducted an expert assessment of item content relevance by recruiting four board-certified neurointensivists who were blinded to patient outcomes and item responses. They reviewed the questions and provided expert opinions regarding the situational relevance of the question construct using the following simple rating scale: likely relevant to some patients, possibly relevant, or unlikely to be relevant. Because the testing was computer adaptive, not all items were used in every patient assessed. We evaluated only items that were used at least 10 times in the cohort. Item responses to symptom frequency and item content relevance ratings were ordinalized (item responses: 5: “very often”, 4: “often”, 3: “sometimes”, 2: “rarely”, 1: “never”; item content relevance: 3: “likely relevant”, 2: “possibly relevant”, and 1: “unlikely to be relevant”). We then assessed for an association between item content relevance and average symptom frequency using Spearman’s rank order correlation to determine whether reported symptom burden was related to item relevance rather than symptom severity.

### Statistical Analyses

We used t-tests and Pearson’s product-moment correlation to assess PROMIS and NIH Toolbox T-Scores. The Wilcoxon rank sum test was used for other continuous or ordinal variables whereas Fisher’s exact test was used for proportions. Exploratory models for successful cognitive testing were constructed separately for patients with ICH and sepsis

including the subset of patients in whom testing was attempted. We included variables with univariate association with successful testing and performed a stepwise variable removal process based on Akaike Information Criteria optimization to address collinearity and overfitting to obtain parsimonious models. Statistical analyses were performed in R version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria).

## Results

We studied 91 critically ill patients, including 53 with sepsis and 38 with ICH. Most patients had no baseline functional impairment (59%); preexisting dementia was rare (8%). At the time of study enrollment, median GCS was 13 [interquartile range 10, 15], the median ICH Score for patients with ICH was 2 [1,3], and the median SOFA was 7 [3, 11] for patients with sepsis, of which 46 (87%) had septic shock. Of the 91 study patients, 9 died in the intensive care unit, 1 declined cognitive testing at the assessment timepoint, and 8 could not be assessed due to feasibility barriers including unanticipated, early discharge and weekend discharge when research staff were unavailable, leaving 73 (80%) patients for assessment.

### Patient-Reported Cognitive Assessments

PROMIS instruments for patient-reported cognitive evaluations were successfully obtained from 42 (58%) of the 73 patients in whom assessments were attempted. The demographic and clinical characteristics of the patients in whom assessments were attempted are detailed in Table 1, stratified by whether a cognitive assessment was obtained successfully or not. Patients who successfully completed a cognitive assessment differed from those who did not by the severity of encephalopathy in the early phase of critical illness (GCS 15 [14, 15] versus 10 [8, 14],  $p<0.001$ ), ICU length of stay (4.5 [2, 6] versus 10 [7.5, 17.5] days,  $p<0.001$ ), the severity of encephalopathy and functional impairment at the time of hospital discharge (GCS 15 [15, 15] versus 10 [8.5, 13.5],  $p<0.001$ ) and modified Rankin Scale 2.5 [1, 4] versus 4 [3, 5],  $p=0.002$ ), the presence of aphasia symptoms (7% versus 42%), and by a premorbid history of dementia (2% versus 19%,  $p=0.042$ ). Half of patients who were able to self-report cognitive symptoms were discharged directly home from the hospital, whereas only 10% of patients unable to complete the assessments went directly home ( $p=0.001$ ). Patient-reported outcomes were obtained successfully in a greater proportion of patients recovering from sepsis than ICH (79% versus 34%,  $p<0.001$ ). We found that simple criteria identified patients able to undergo successful patient-reported outcome assessment: GCS 15 at the time of hospital discharge as a single criterion was 92% sensitive, 79% specific and 85% accurate, and the combined criteria of GCS 15 and ICU length of stay less than one week was 97% sensitive, 71% specific and 82% accurate. Notably, 34 of 37 (92%) patients with discharge GCS scores of 15 but only 1 of 6 (17%) patients with discharge GCS scores of 14 were able to complete any PROMIS assessment.

PROMIS Cognition General Concerns scores were about two standard deviations *better* than the reference population mean ( $31 \pm 12$ ,  $p<0.001$ , corresponding to the 97<sup>th</sup> [93<sup>rd</sup>, 99.9<sup>th</sup>] percentile), although cognition-related activity impairments, as measured by PROMIS Cognition Abilities, were not different than the reference population ( $50 \pm 9$ ,  $p=0.9$ ). Cognitive testing results for PROMIS instruments and NIH Toolbox tests are summarized in

Table 2. There was no difference in PROMIS cognitive scores between 8 patients with ICH and 30 with sepsis.

### Objective Cognitive Assessments

Only 9 (12%) patients in whom assessments were attempted were able to complete any of the three NIH Toolbox cognitive tests. All patients who were able to perform an NIH Toolbox cognitive test scored at the maximum (15) on the GCS at the time of assessment, were able to complete PROMIS assessments and did not differ from untestable patients by self-reported cognitive symptoms (mean PROMIS Cognition General Concerns 30 versus 31,  $p=0.8$ ). Objective cognitive tests were obtained successfully in a greater proportion of patients recovering from sepsis than ICH (21% versus 3%,  $p<0.001$ ). Performance on individual instruments showed some evidence, although not statistically significant in this sample, for impairment in working memory (List Sorting Working Memory test mean  $42 \pm 11$ ,  $p=0.06$  compared to demographically-adjusted population), along with severe impairments in cognitive processing speed (Pattern Comparison Processing Speed test mean  $28 \pm 9$ ,  $p<0.001$ ), attention and executive function (Flanker Inhibitory Control and Attention test mean  $33 \pm 4$ ,  $p<0.001$ ). The mean of the three cognitive test scores was  $34 \pm 6$  ( $p<0.001$  compared to population), corresponding to the 5<sup>th</sup> [2<sup>nd</sup>, 9<sup>th</sup>] percentile. Only one patient with ICH was able to complete the Toolbox testing, scoring 38 for working memory, 28 for processing speed and 37 for attention and executive function, which is very close to the mean for the sepsis patients.

### Comparison of Objective and Patient-Reported Cognition

PROMIS Cognition General Concerns scores were negatively correlated with Flanker Inhibitory Control and Attention scores ( $\rho = -0.81$ ,  $p=0.026$ ), which is consistent with the scoring system in which higher scores on the PROMIS instrument and lower scores on the Flanker test indicate worse function, with a caveat that the same size was small ( $n=7$ ). There was no significant correlation between PROMIS Cognition General Concerns and the List Sorting Working Memory Test ( $\rho = -0.07$ ,  $p=0.85$ ) or the Pattern Comparison Processing Speed Test ( $\rho = -0.15$ ,  $p=0.73$ ). Patients who completed objective testing scored a median 88 [65, 94] percentile ranks lower than their self-reported cognitive status.

### Instrument Failure Analysis: Non-Completion and Test Stopping

The low rate of response for the objective and patient-reported assessments was not anticipated during the study design, so our data collection system did not have a systematic way of prospectively adjudicating the cause of assessment failure. Qualitative review of free text notes and study team member feedback identified inattention, embarrassment related to performance difficulty during the exam and frustration as factors in non-completion of the objective assessments. Non-completion of the patient-reported assessment was also influenced by those factors as well as confusion about how to interpret and answer some instrument questions. The exploratory analysis of testing barriers in patients with ICH found that lower ICH Score (i.e. lower ICH severity; OR 0.056 per increase in ICH Score, 95% CI [0.002, 0.36],  $p=0.018$ ) and shorter ICU length of stay (OR 0.61 per day [0.35, 0.86],  $p=0.024$ ) were the strongest independent predictors of successful cognitive testing. In patients with sepsis, shorter ICH length of stay was also an independent predictor of

successful cognitive testing (OR 0.74 [0.53, 0.91],  $p=0.022$ ), whereas a premorbid history of dementia reduced the success of cognitive testing (OR 0.054 [0.002, 0.66],  $p=0.035$ ). As noted in the methods, the stopping rule algorithm for the PROMIS computer adaptive testing constrains the test from four to 12 items, stopping the test after 12 questions even if the goal estimated precision of the score measurement is not reached. In our sample, 29% of tests were stopped by reaching the maximum 12 items, and in 83% of those cases the patients were scored at the floor (T score 14.5).

### **Instrument Failure Analysis: Expert Assessment of Item Content Relevance**

Fifteen of the 34 items in the PROMIS General Concerns assessment item bank were used in at least 10 individuals, and those 15 items represented 260 (94%) of the 276 individual item administrations. Among those 15 items, patients reported never experiencing a single instance of the cognitive impairment symptom described in at least 90% of responses to 8 of the 15 questions. Results for each item are summarized in the Online Supplemental Table. For example, 100% of patients responded “never” to the items “I have made mistakes when writing down phone numbers” and “I have had trouble finding my way to a familiar place”, and all four experts rated those questions as unlikely to be situationally relevant. In contrast, the majority of patients endorsed experiencing problems in the last week in response to the statements like “I have had to work harder than usual to keep track of what I was doing”, “My thinking has been slow” and “It has seemed like my brain was not working as well as usual”, which received high situational relevance ratings from experts. Scores for individual items were strongly correlated with experts’ independent assessment of the item’s situational relevance ( $\rho = 0.72$ ,  $p=0.002$ ).

## **Discussion**

We observed that many patients in the early stage of recovery from critical illness, especially those with neurologic injury, were unable to self-report their cognitive status. Moreover, a large majority of them were unable to engage with objective cognitive testing at the bedside, even when able to engage the examiner sufficiently to answer questions for self-reporting. Among patients able to complete objective tests, all of whom scored normal by GCS assessment, the tested domains of cognitive function were impaired, especially executive function. Scores from the self-reported cognitive assessment, interpreted at face value, paradoxically indicated a low burden of cognitive impairment symptoms, although the burden of reported symptoms was strongly correlated to the situational relevance of the symptoms described according to expert ratings. Thus, symptom severity measurement by these PROs was more indicative of item relevance than true symptom burden, indicating poor construct validity. Our study was designed under the assumption that these cognitive assessment methods, which had already been validated in multiple populations of patients with chronic medical and neurologic diseases, successfully implemented by our research team in patients with chronic metabolic encephalopathy, and incorporated into the design of a major intensive care outcomes study, would adapt well to the post-critical care population after resolution of critically illness.<sup>25</sup> A critical appraisal of our methods and results against the background of related research offers insights to explain our findings and may guide the development of better measurement methods.

Although patient-reported outcomes need not necessarily yield results that are highly correlated with objective measurements, the severe discordance in patient-reported and objective findings here suggests that the construct of the patient-reported assessments were not suitable for this population. Patient-reported outcome (PRO) instruments use a set of questions (the item bank) designed to discriminate patients according to function or symptom severity. Due to individual variation, not every question will be relevant to the experiences and current life situation of every patient, but as long as most questions are relevant, the instrument should provide a score within a tolerable degree of precision. We chose a CAT implementation because the adaptive algorithm will increase the length of the test when answers to the first few questions are discordant in order to achieve good precision, which can be useful when the discrimination value of some questions may be reduced.<sup>23</sup> However, CAT algorithms cannot improve measurements when too many questions in the item bank are situationally irrelevant and answers cause paradoxical discrimination scores. For example, in the general population, answering “never” to the question “Within the last week, I have had trouble finding my way to a familiar place” would indicate better cognition, but when the reason for the answer is that you haven’t gone anywhere by yourself for a week because you’ve been in a medical institution, the score interpretation becomes a misdirection. The strong relationship between the situational relevance of the symptoms described in the item statements and the observed responses suggests that many item constructs were not applicable for the context of hospitalization. By analogy, the item “I am able to walk 10 blocks without assistance” in a PRO measuring physical fitness may show good discrimination performance in the general population, but would paradoxically score highly fit paraplegic athletes as having low fitness, and a physical fitness PRO utilizing many items related to ambulatory functions would have poor construct validity in that population. The NIH Toolbox misclassified many verbally responsive, interactive patients as unable to be assessed, inconsistent with data from the GCS and PROMIS assessments. Again, this is likely due to test design. Although the NIH Toolbox was created to reduce barriers to feasibility to psychometric testing in research, cognitive testing using NIH Toolbox instruments was still infeasible for most patients in early illness recovery with many patient unable to perform the tests due to psychomotor delay, inattention, trouble processing instructions or frustration.

Previous studies evaluating post-intensive care cognitive symptoms have performed the assessment several months or more after hospital discharge as intermediate or long-term outcome measures, using objective neuropsychological tests like the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) or patient-reported instruments.<sup>4,27–30</sup> Similar to the NIH Toolbox, the RBANS has a simplified design enabling bedside assessments, and resources have recently become available to calculate normative scores adjusted for age, gender and education, although the test-retest stability characteristics of NIH Toolbox cognitive tests are uniformly superior and more in line with ideal values for clinical use.<sup>15,31,32</sup> One study has proposed to use the RBANS for early cognitive assessment after critical illness at the identical timepoint we tested here, but the published study protocol provides no preliminary data or cited literature to establish feasibility.<sup>6</sup> Initial results from that study show findings congruent with our results: after excluding patients who were expected to face barriers to successful assessments (e.g. premorbid cognitive



impairment, neurological injury, unlikely to adhere with follow-up), only 50% of approached subjects were recruited, and early cognitive assessments were unobtainable in 35%.<sup>33</sup> We have recently shown that the NIH Toolbox cognitive battery is feasible and valid in medically ill patients with advanced, chronic liver failure, but cognitive performance was less impaired in that sample.<sup>25</sup>

An abbreviated version of the Cognitive Failures Questionnaire (CFQ) was recently adapted from the full questionnaire based on sampling between 12 and 24 months after ICU discharge. Like objective cognitive tests, the self-reported CFQ has not been evaluated in patients in the early course of recovery, and many of the items would likely have limited situational relevance to patients in the hospital, rehabilitation facilities, skilled nursing facilities and other supervised care environments (e.g. “Do you find you forget why you went from one part of the house to the other?”, “Do you fail to notice signposts on the road?”, “Do you find you forget which way to turn on a road you know well but rarely use?”), similar to what we observed with the PROMIS items.<sup>30,34</sup> A recent study used the abbreviated CFQ to compare patient-reported and objective cognitive function in ICU survivors from 3-12 months after discharge similarly found no correlation between the two measurement approaches.<sup>35</sup> Their analysis did not explore the performance or content validity of individual items in the abbreviated CFQ, but did make an interesting observation that the severity of patient’s self-reported cognitive symptoms was well correlated with the severity of their self-reported symptoms of post-traumatic stress, anxiety and depression, which raises psychological symptom comorbidity as another confounder alongside item content.

Ceiling and floor effects are important psychometric test properties when considering instruments to assess subjects over a wide range of abilities. Although GCS measurements are prognostically useful in critically ill patients, approximately 40% of critically ill patients never score below maximum on the GCS during the initial 24 hour resuscitation and stabilization period, with increasing compression of scores at the ceiling as physiologic stability improves.<sup>14</sup> Delirium instruments are limited by both floor and ceiling effects. Major studies utilizing CAM-ICU found that 13-18% of subjects could not be assessed due to communication impairments (other than intubation), another 14-19% were not assessable due to persistent coma, and 63% of subjects were comatose for at least part of the ICU stay.<sup>1</sup> The NIH Toolbox cognition measures did not show ceiling or floor effects during its derivation and validation, but in this sample of severely compromised patients there was a clear clustering at the floor of the reference distribution. The difference in success rates for testing patients after critical care for ICH and sepsis confirms that testing barriers including language, motor and attentional impairments can vary according to diagnosis.

There are important limitations to these data. This is a single-center sample that was limited to two specific conditions. Although restricting inclusion to ICH as a representative neurologic illness and sepsis as a representative systemic illness was helpful in comparing and contrasting patients with direct and indirect brain injury, the characteristics of these groups may not generalize well to other patient populations. Modifications to the PROMIS items or NIH Toolbox testing methods may have enabled successful testing of more patients but would have diminished the standardization and interpretability of the data. Addition of

delirium assessments could have enhanced our understanding of testing barriers. Delirium is conceptualized as a cognitive impairment syndrome with a fluctuating course. In addition to a fluctuating course, for example, the Diagnostic and Statistical Manual of Mental Disorders, fifth edition, criteria require a “disturbance in attention and awareness” and “at least one additional disturbance in cognition. The cross-sectional cognitive assessments we performed do not enable us to determine whether patients’ cognitive symptoms were fluctuating, but it is likely that some were experiencing delirium.<sup>36</sup> The observed rate of assessment non-completion was higher than anticipated, and we did not design our study database to prospectively ascertain the reasons for assessment failure. That factor limited our ability to systematically determine the cause of assessment failure beyond describing the most frequent causes. The sample size of patients with ICH and use of summary descriptors (e.g. ICH Score, hematoma volume, hemorrhage location) in exploratory factor analysis limits our ability to analyze for potentially relevant, domain-specific impairments. Concurrent testing with additional objective and patient-reported instruments would have allowed us to compare their performance, but the construct and content of alternative tests like RBANS and CFQ are similar to the tests we used and we wished to mitigate the risk of fatigue and inattention further confounding our data. Finally, use of narrow inclusion and exclusion criteria to preselect for patients with the most preserved neurologic function would have yielded greater response rates, but the ability threshold of this population to use these instruments was not prospectively known and the patients with greatest risk factors for impairment are the most clinically relevant population to study.

Overcoming the feasibility and validity barriers to assessing cognitive performance after critical illness may require development of new patient-reported and objective tests adapted to the abilities of the population and ecologically valid for the healthcare environment, analogous to how cognitive tests have been developed for young children. Patients could be screened for sufficient cognitive capacity and absence of delirium to render testing feasible using simple bedside measures like GCS score and Confusion Assessment Method. For patient-reported assessments, creating an item bank of questions that are situationally relevant is imperative for the items to correctly discriminate by severity. Modifying the style of questions (e.g. framing answers as yes/no rather than multiple choice) may overcome barriers related to inattention and slow processing. Objective measurements could be improved by less dependence on task completion speed as a surrogate for ability, which is especially problematic for test that require the patient to use fine motor skills, and instruments need to be calibrated to a lower average level of function to avoid floor effects. Ideally, new instruments could be developed for high impairments patients with design characteristics that optimize their validity, and crosswalked to scores on instruments that perform well in the general population so that patients can be followed over the course of their acute illness and recovery. Methods for assessing cognition during early critical illness recovery could be further informed by additional research employing mixed methods that are designed to specifically compare the strengths and limitations of the wide variety of assessment methods currently available in different patient groups.

## Conclusions

Methods of objective and patient-reported cognitive testing that have been validated for use in patients with chronic medical and neurologic illness were infeasible or yielded invalid results among patients in this study who were in early recovery from critical illness. Longer critical illness duration and worse neurocognitive impairments, whether chronic or acute, reduced testing feasibility.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding:** Primary funding for this study and support for Dr. Maas and Dr. Griffith came from National Institutes of Health (NIH) grant K23NS092975. Dr. Maas received additional related support from NIH grant L30NS080176. Dr. Griffith received additional support through R01MD010440. Dr. Reid received support through the Northwestern Center for Circadian and Sleep Medicine. Drs. Reid and Zee receive support from NIH grants UM1HL112856, R01HL140580 and P01AG011412. Research reported in this publication was supported, in part, by the NIH grant UL1TR000150. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**Conflicts of Interest:** Dr. Maas reports grants from the National Institutes of Health and Northwestern Memorial Foundation during the conduct of the study. Dr. Liotta reports a grant from the National Institutes of Health during the conduct of the study. Drs. Reid and Zee reports grants from the National Institutes of Health and the Defense Advanced Research Projects Agency during the conduct of the study. Dr. Griffith reports grants from the National Institutes of Health during the conduct of the study.

## References

1. Maas MB, Naidech AM. Critical Care Neurology Perspective on Delirium. *Semin Neurol* 2016;36(6):601–606. (In eng). DOI: 10.1055/s-0036-1592318. [PubMed: 27907964]
2. Ehlenbach WJ, Hough CL, Crane PK, et al. Association between acute care and critical illness hospitalization and cognitive function in older adults. *JAMA* 2010;303(8):763–70. (In eng). DOI: 10.1001/jama.2010.167. [PubMed: 20179286]
3. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA* 2004;291(14):1753–62. (In eng). DOI: 10.1001/jama.291.14.1753. [PubMed: 15082703]
4. Pandharipande PP, Girard TD, Jackson JC, et al. Long-term cognitive impairment after critical illness. *N Engl J Med* 2013;369(14):1306–16. (In eng). DOI: 10.1056/NEJMoa1301372. [PubMed: 24088092]
5. Needham DM, Davidson J, Cohen H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. *Crit Care Med* 2012;40(2):502–9. (In eng). DOI: 10.1097/CCM.0b013e318232da75. [PubMed: 21946660]
6. Wilcox ME, Lim AS, McAndrews MP, et al. A study protocol for an observational cohort investigating COGNitive outcomes and WELLness in survivors of critical illness: the COGWELL study. *BMJ Open* 2017;7(7):e015600. (In eng). DOI: 10.1136/bmjopen-2016-015600.
7. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974;2(7872):81–4. (In eng). DOI: 10.1016/s0140-6736(74)91639-0. [PubMed: 4136544]
8. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22(7):707–10. (In eng). DOI: 10.1007/bf01709751. [PubMed: 8844239]

9. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315(8):801–10. (In eng). DOI: 10.1001/jama.2016.0287. [PubMed: 26903338]
10. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006;34(5):1297–310. (In eng). DOI: 10.1097/01.CCM.0000215112.84523.F0. [PubMed: 16540951]
11. Schmidt FA, Liotta EM, Prabhakaran S, Naidech AM, Maas MB. Assessment and comparison of the max-ICH score and ICH score by external validation. *Neurology* 2018;91(10):e939–e946. (In eng). DOI: 10.1212/WNL.0000000000006117. [PubMed: 30068631]
12. Maas MB, Rosenberg NF, Kosteva AR, et al. Surveillance neuroimaging and neurologic examinations affect care for intracerebral hemorrhage. *Neurology* 2013;81(2):107–12. (In eng). DOI: 10.1212/WNL.0b013e31829a33e4. [PubMed: 23739227]
13. Maas MB, Berman MD, Guth JC, Liotta EM, Prabhakaran S, Naidech AM. Neurochecks as a Biomarker of the Temporal Profile and Clinical Impact of Neurologic Changes after Intracerebral Hemorrhage. *J Stroke Cerebrovasc Dis* 2015;24(9):2026–31. (In eng). DOI: 10.1016/j.jstrokecerebrovasdis.2015.04.045. [PubMed: 26143415]
14. Knox DB, Lanspa MJ, Pratt CM, Kuttler KG, Jones JP, Brown SM. Glasgow Coma Scale score dominates the association between admission Sequential Organ Failure Assessment score and 30-day mortality in a mixed intensive care unit population. *J Crit Care* 2014;29(5):780–5. (In eng). DOI: 10.1016/j.jcrc.2014.05.009. [PubMed: 25012961]
15. Weintraub S, Dikmen SS, Heaton RK, et al. Cognition assessment using the NIH Toolbox. *Neurology* 2013;80(11 Suppl 3):S54–64. (In eng). DOI: 10.1212/WNL.0b013e3182872ded. [PubMed: 23479546]
16. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63(11):1179–94. (In eng). DOI: 10.1016/j.jclinepi.2010.04.011. [PubMed: 20685078]
17. Fieo R, Ocepek-Welikson K, Kleinman M, et al. Measurement Equivalence of the Patient Reported Outcomes Measurement Information System. *Psychol Test Assess Model* 2016;58(2):255–307. (In eng). [PubMed: 28523238]
18. Maas MB, Lizza BD, Abbott SM, et al. Factors Disrupting Melatonin Secretion Rhythms during Critical Illness. *Crit Care Med* 2020.
19. Maas MB, Lizza BD, Kim M, et al. Stress-Induced Behavioral Quiescence and Rest-Activity Dysrhythmia during Critical Illness. *Crit Care Med* 2020.
20. Rost NS, Smith EE, Chang Y, et al. Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. *Stroke* 2008;39(8):2304–9. (In eng). DOI: 10.1161/STROKEAHA.107.512202. [PubMed: 18556582]
21. Maas MB, Francis BA, Sangha RS, Lizza BD, Liotta EM, Naidech AM. Refining Prognosis for Intracerebral Hemorrhage by Early Reassessment. *Cerebrovasc Dis* 2017;43(3-4):110–116. (In eng). DOI: 10.1159/000452679. [PubMed: 28049196]
22. Wilson JT, Hareendran A, Grant M, et al. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. *Stroke* 2002;33(9):2243–6. (In eng). DOI: 10.1161/01.str.0000027437.22450.bd. [PubMed: 12215594]
23. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16 Suppl 1:133–41. (In eng). DOI: 10.1007/s11136-007-9204-6. [PubMed: 17401637]
24. Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual Life Res* 2010;19(1):125–36. (In eng). DOI: 10.1007/s11136-009-9560-5. [PubMed: 19941077]
25. Kim M, Liotta EM, Zee PC, et al. Impaired cognition predicts the risk of hospitalization and death in cirrhosis. *Ann Clin Transl Neurol* 2019;6(11):2282–2290. (In eng). DOI: 10.1002/acn3.50924. [PubMed: 31631586]

26. Cella D, Choi S, Garcia S, et al. Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Qual Life Res* 2014;23(10):2651–61. (In eng). DOI: 10.1007/s11136-014-0732-6. [PubMed: 24938431]
27. Needham DM, Dinglas VD, Morris PE, et al. Physical and cognitive performance of patients with acute lung injury 1 year after initial trophic versus full enteral feeding. EDEN trial follow-up. *Am J Respir Crit Care Med* 2013;188(5):567–76. (In eng). DOI: 10.1164/rccm.201304-0651OC. [PubMed: 23805899]
28. Davydow DS, Zatzick D, Hough CL, Katon WJ. In-hospital acute stress symptoms are associated with impairment in cognition 1 year after intensive care unit admission. *Ann Am Thorac Soc* 2013;10(5):450–7. (In eng). DOI: 10.1513/AnnalsATS.201303-060OC. [PubMed: 23987665]
29. Marra A, Pandharipande PP, Girard TD, et al. Co-Occurrence of Post-Intensive Care Syndrome Problems Among 406 Survivors of Critical Illness. *Crit Care Med* 2018;46(9):1393–1401. (In eng). DOI: 10.1097/CCM.0000000000003218. [PubMed: 29787415]
30. Wassenaar A, de Reus J, Donders ART, et al. Development and Validation of an Abbreviated Questionnaire to Easily Measure Cognitive Failure in ICU Survivors: A Multicenter Study. *Crit Care Med* 2018;46(1):79–84. (In eng). DOI: 10.1097/CCM.0000000000002806. [PubMed: 29068855]
31. Olaithe M, Weinborn M, Lowndes T, et al. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Normative Data for Older Adults. *Arch Clin Neuropsychol* 2019 (In eng). DOI: 10.1093/arclin/acy102.
32. Duff K, Beglinger LJ, Schoenberg MR, et al. Test-retest stability and practice effects of the RBANS in a community dwelling elderly sample. *J Clin Exp Neuropsychol* 2005;27(5):565–75. (In eng). DOI: 10.1080/13803390490918363. [PubMed: 16019633]
33. Wilcox ME, McAndrews MP, Van J, et al. Sleep Fragmentation and Cognitive Trajectories after Critical Illness. *Chest* 2020 (In eng). DOI: 10.1016/j.chest.2020.07.036.
34. Broadbent DE, Cooper PF, FitzGerald P, Parkes KR. The Cognitive Failures Questionnaire (CFQ) and its correlates. *Br J Clin Psychol* 1982;21 (Pt 1):1–16. (In eng). [PubMed: 7126941]
35. Brück E, Larsson JW, Lasselin J, et al. Lack of clinically relevant correlation between subjective and objective cognitive function in ICU survivors: a prospective 12-month follow-up study. *Crit Care* 2019;23(1):253. (In eng). DOI: 10.1186/s13054-019-2527-1. [PubMed: 31300016]
36. American Psychiatric Association. *Neurocognitive Disorders. Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.

**Details Page**

This manuscript complies with all instructions to authors.

The authorship requirements have been met and the final manuscript was approved by all authors.

We confirm that this study adhered to ethical guidelines, was conducted under IRB approval and informed consent was obtained from all participants.

The STROBE reporting checklist was used.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1:**

## Demographic and Clinical Characteristics of Patients Undergoing Early Post-Critical Illness Assessments

Baseline Characteristics	No Successful Testing	Any Successful Testing	p
Number of patients	31	42	
Type = sepsis (%)	8 (25.8)	30 (71.4)	<0.001
Age (mean (SD))	68.74 (12.62)	65.38 (17.41)	0.37
Male (%)	19 (61.3)	26 (61.9)	1
Race (%)			0.42
Asian	2 ( 6.5)	1 ( 2.4)	
Black	11 (35.5)	11 (26.2)	
White	18 (58.1)	30 (71.4)	
Ethnicity (%)			0.69
Hispanic or Latino	4 (12.9)	5 (11.9)	
Not Hispanic or Latino	27 (87.1)	36 (85.7)	
Unknown or Not Reported	0 ( 0.0)	1 ( 2.4)	
Premorbid dementia (%)	6 (19.4)	1 ( 2.4)	0.042
Premorbid disability by mRS (%)			0.35
0: Asymptomatic	17 (56.7)	23 (59.0)	
1: Symptoms with disability	1 ( 3.3)	3 ( 7.7)	
2: Slight disability but independent	4 (13.3)	2 ( 5.1)	
3: Moderate disability	4 (13.3)	5 (12.8)	
4: Moderately severe disability	2 ( 6.7)	6 (15.4)	
5: Severe disability	2 ( 6.7)	0 ( 0.0)	
<b>Clinical Characteristics during First Day of ICU Care</b>			
ICH Score (median [IQR])	2.00 [2.00, 3.00]	1.00 [0.00, 4.00]	0.13
ICH Location			0.42
Lobar	10	3	
Deep	12	9	
Cerebellum	3	0	
Brainstem	1	0	
ICH volume (in mL; median [IQR])	29.3 [18.8, 46.4]	10.1 [5.5, 24.8]	0.057
Sepsis severity = Septic shock (%)	6 (75.0)	28 (93.3)	0.39
Median GCS (median [IQR])	10.00 [8.00, 14.00]	15.00 [14.00, 15.00]	<0.001
Median RASS (median [IQR])	-2.00 [-3.50, 0.00]	0.00 [0.00, 0.00]	<0.001
SOFA (median [IQR])	5.00 [3.00, 8.50]	4.00 [1.50, 8.75]	0.36
Mechanical ventilation (%)	17 (54.8)	12 (28.6)	0.043
Intravenous sedation (%)	15 (48.4)	15 (35.7)	0.40
Aphasia symptoms (%)	13 (41.9)	3 (7.1)	0.001
<b>Discharge Status</b>			
ICU length of stay (median [IQR] days)	10.00 [7.50, 17.50]	4.50 [2.00, 6.00]	<0.001

Baseline Characteristics	No Successful Testing	Any Successful Testing	p
Hospital disposition (%)			0.001
Dead	3 (9.7)	1 (2.4)	
Home	3 (9.7)	21 (50.0)	
Institution	25 (80.6)	20 (47.6)	
Discharge GCS (median [IQR])	10.00 [8.50, 13.50]	15.00 [15.00, 15.00]	<0.001
Discharge mRS (median [IQR])	4.00 [3.00, 5.00]	2.50 [1.00, 4.00]	0.002

mRS: modified Rankin Scale, GCS: Glasgow Coma Scale, RASS: Richmond Agitation Sedation Scale, SOFA: Sequential Organ Failure Assessment

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2:**

## Cognitive Testing Results

Cognitive Test	Result	P-value*
Number of Patients Completing PROMIS	42	
PROMIS Cognition - Abilities (mean (SD))	50.1 (8.5)	0.9
PROMIS Cognition - General (mean (SD))	30.8 (12.4)	<0.001
Number of Patients Completing NIH Toolbox	9	
List Sorting Working Memory (mean (SD))	41.9 (11.1)	0.06
Pattern Comparison Processing Speed (mean (SD))	27.9 (9.1)	<0.001
Flanker Attention and Inhibitory Control (mean (SD))	33.4 (3.9)	<0.001
Average of NIH Toolbox tests (mean (SD))	34.0 (6.3)	<0.001

\*P-values are compared to general population.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript