Brain Injury Functional Outcome Measure (BI-FOM): A Single Instrument Capturing the Range of Recovery in Moderate-Severe Traumatic Brain Injury

John Whyte, MD, PhD[a], **Joseph T. Giacino, PhD**[b], **Allen W. Heinemann, PhD**[c,d], **Yelena Bodien, PhD**[b], **Tessa Hart, PhD**[a], **Mark Sherer, PhD**[e], **Gale G. Whiteneck, PhD**[f], **David Mellick, MA**[f], **Flora M. Hammond, MD**[g], **Patrick Semik, BA**[d], **Amy Rosenbaum, PhD**[h], **Risa Nakase Richardson, PhD**[i,j]

[a]Moss Rehabilitation Research Institute, Albert Einstein Healthcare Network, Elkins Park, Pennsylvnia;

[b]Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, Massachusetts;

[c]Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, Illinois;

[d]Center for Rehabilitation Outcomes Research, Shirley Ryan Ability Lab, Chicago, Illinois;

[e]TIRR Memorial Hermann, Houston, Texas;

[f]Craig Hospital Research Department, Englewood, Colorado;

[g]Department of Physical Medicine and Rehabilitation, Indiana University School of Medicine and Rehabilitation Hospital of Indiana, Indianapolis, Indiana;

[h]Park Terrace Care Center, Flushing, New York;

[i]MHBS, DVBIC, James A. Haley Veterans Hospital, Tampa, Florida;

[j]Division of Pulmonary and Sleep Medicine, Department of Internal Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida.

Abstract

Objective: To develop a measure of global functioning after moderate-severe TBI with similar measurement precision but a longer measurement range than the FIM.

Design: Phase 1: retrospective analysis of 5 data sets containing FIM, Disability Rating Scale, and other assessment items to identify candidate items for extending the measurement range of the FIM; Phase 2: prospective administration of 49 candidate items from phase 1, with Rasch analysis to identify a unidimensional scale with an extended range.

Setting: Six TBI Model System rehabilitation hospitals.

Corresponding author: John Whyte, MD, PhD, Moss Rehabilitation Research Institute, 50 Township Line Rd, Philadelphia, PA 19027. jwhyte@einstein.edu.

Participants: Individuals (NZ184) with moderate-severe injury recruited during inpatient rehabilitation or at 1-year telephone follow-up. Interventions: Participants were administered the 49 assessment items in person or via telephone.

Main Outcome Measures: Item response theory parameters: item monotonicity, infit/outfit statistics, and Factor 1 variance.

Results: After collapsing misordered rating categories and removing misfitting items, we derived the Brain Injury Functional Outcome Measure (BI-FOM), a 31-item assessment instrument with high reliability, greatly extended measurement range, and improved unidimensionality compared with the FIM.

Conclusions: The BI-FOM improves global measurement of function after moderate-severe brain injury. Its high precision, relative lack of floor and ceiling effects, and feasibility for telephone follow-up, if replicated in an independent sample, are substantial advantages.

Severe traumatic brain injury (TBI) results in a period of coma,[1] after which the trajectory of recovery is protracted and unpredictable.[2] Patients may remain unconscious and transition to a vegetative state, and they may pass through stages of minimal consciousness and/or posttraumatic confusion, with most gradually attaining greater functional independence. A similar pattern may occur on an accelerated trajectory with moderate TBI. Measurable recovery may continue for years after moderate-severe injury.[3]

Measuring the functional outcomes of moderate-severe TBI is challenging from both conceptual and practical standpoints. TBI affects many domains—physical, cognitive, behavioral, psychosocial—in varying patterns, and recovery is heterogeneous, with some patients remaining unconscious for prolonged periods and others eventually resuming full functioning. Accordingly, the National Institute on Neurological Disorders and Stroke Common Data Elements initiative for TBI endorsed multiple outcome measures to cover the important domains and levels of function.[4] Such focused instruments are useful tor measuring specific domains or severity strata and as outcomes for focused interventions. Nonetheless, a single measure of global function related to diffuse and/or multifocal neurologic impairment throughout the course of recovery from moderate-severe TBI would allow for quantitative characterization of the recovery trajectory, better understanding of the effect of demographic and clinical variables on that trajectory, and improved assessment of interventions expected to affect global functioning, such as early resuscitation protocols or neuroprotective agents. Furthermore, early measures of functional status are among the strongest predictors of later outcome and, thus, are useful severity indicators in both observational research and clinical trials.

Available measures are not sufficient for either long-term measurement or early severity adjustment. The Disability Rating Scale,[5] which measures the range "from coma to community," is ordinal rather than interval. It thus lacks precision, with particularly wide spacing at the upper end, making it insensitive to functionally important changes and

problematic for parametric analyses.[6] The widely used FIM[7] allows interval scoring but exhibits both floor and ceiling effects when used to track the progress of people with TBI over the first year or longer.[8,9] Thus, it obscures meaningful differences in injury severity early after injury and meaningful differences in higher level functioning at later time points. The Coma Recovery Scale—Revised (CRS-R),[10] a standardized measure designed to detect behavioral signs of consciousness, avoids the FIM's floor effects but does not measure improvements beyond emergence from the minimally conscious state. The Glasgow Outcome Scale—Extended,[11] while an improvement over the original Glasgow Outcome Scale, is also ordinal and relatively insensitive to change. Psychometric batteries are sensitive at higher but not lower levels of function and are labor intensive, typically requiring in-person administration.[12]

The ideal measure for this purpose would have (1) measurement sensitivity throughout the range of injury severity and long-term recovery, (2) unidimensional interval measurement of the construct of global functioning,[13] and (3) the capacity to be administered by telephone.

The goal of this study was to develop a unidimensional measure of global function centered on, and with similar measurement precision to, the FIM but with a lower floor and a higher ceiling (ie, a "longer" scale). In designing the study, we faced a conceptual challenge. The functional abilities assessed by the FIM are substantially normative, in that almost everyone without functional limitations performs activities such as dressing and locomotion independently. Similarly, items capable of assessing extremely impaired functioning are also normative (eg, keeping the eyes open, breathing independently). In contrast, as functional abilities increase above those assessed by the FIM, this increased capacity is shown in varied ways. Some people with strong functional abilities run businesses and some coach baseball, but neither activity could measure normative function because they reflect individual preferences. Thus, we expected greater challenges in raising the ceiling (while retaining a unitary dimension of "global neurologic function") than in lowering the floor. However, we anticipated that items measuring neuropsychological functions commonly affected by TBI such as attention, memory, and executive function might be suitable for the purpose of extending the ceiling of the new measure. For the reasons described above, we sought measures of such abilities that could be validly assessed via telephone.

Methods

The study was conducted in 2 phases by 6 centers participating in the National Institute on Disability Independent Living and Rehabilitation Research Traumatic Brain Injury Model System (TBIMS) program.

Retrospective phase: identifying potential items to include in the Brain Injury Functional Outcome Measure

Through discussion with project team members, we located 5 deidentified data sets in which the FIM and/or DRS had been administered to a sample of participants with moderate-severe TBI, along with 1 or more other measures that sampled a broad range of functioning. These included data from the TBIMS National Database: (1) a current data set with a telephone-administered cognitive battery, the Brief Test of Adult Cognition by Telephone[14]; (2) an

archived data set that included in-person neuropsychological tests; (3) a placebo-controlled trial of amantadine hydrochloride in patients with disorders of consciousness (DOC)[15]; (4) a study on assessment of the posttraumatic confusional state[16]; and (5) a clinical database from a program serving individuals with DOC.[17] The posttraumatic confusional state study participants were also enrolled in the TBIMS, allowing their cognitive data to be linked with their FIM and DRS data. The items contained in these data sets are shown in supplemental table S1 (available online only at http://www.archives-pmr.org/). Altogether we examined the DRS, CRS-R, Cognitive Test for Delirium,[18,19] Toronto Test of Acute Recovery after TBI,[20] and Mississippi Aphasia Screening Test[21] for items that might extend the floor, and we examined the Supervision Rating Scale,[22] various neuropsychological measures (both in-person and telephone-administered), and the Participation Assessment with Recombined Tools—Objective,[23] for items that might extend the ceiling.

We used item response theory analysis to cocalibrate the items with FIM and/or DRS to establish their dimensionality. Candidate items were dropped if they misfit the global functional dimension represented by the FIM and DRS or were redundant in difficulty level with existing items. Using this approach, we arrived at a set of 49 items for prospective administration, as shown in supplemental table S2 (available online only at http://www.archives-pmr.org/). As detailed below, these items were then administered prospectively to participants with moderate-severe TBI (see fig 1 for an overview of the study design).

Prospective phase: calibration of the Brain Injury Functional Outcome Measure items

Participants—Participants in this phase were 184 individuals with moderate-severe TBI enrolled at the 6 participating TBIMS sites. They were assessed either as inpatients, shortly after discharge, or at approximately 1 year post injury, depending on the capacities of each participating system. All participants met TBIMS eligibility criteria[24] in that they were 16 years or older, met diagnostic criteria for moderate-severe TBI, and received emergency and/or acute care within 72 hours of injury followed by transfer to a TBIMS-affiliated inpatient rehabilitation unit within 72 hours of acute care. Inpatients were assessed in person, while those who had been discharged were assessed via telephone interview with the individual or a caregiver. Sixteen participants met all TBIMS criteria except that they received acute care at a non-TBIMS-affiliated hospital; for this reason, some of their demographic and early clinical data are missing. We included these participants to obtain greater representation of patients with the most severe injuries. Participants were excluded if they had a sensory, cognitive, or motor disability prior to their TBI, were not fluent in English prior to injury, or were experiencing an acute illness that might depress function. Informed consent was obtained from the participant or a legally authorized representative, and the local Institutional Review Board at each site approved the study. Demographic and clinical characteristics are summarized in table 1. The distribution of assessment time post injury was bimodal, as shown in fig 2, reflecting the preponderance of assessments early after injury or at about a year.

Procedure—Inpatient assessment began with the CRS-R. If the patient scored at least in the minimally conscious state+ range,[25] the remaining cognitive items were administered

in person, in ascending order of difficulty as estimated from the retrospective calibration. Patients assessed by phone were given the highest possible CRS-R score if they could participate meaningfully in the remaining telephone data collection. Administration and scoring guidelines were standardized and administered according to available instructions, except that the instructions for the Conceptual Reasoning and Cognitive Test for Delirium Vigilance measures were modified for telephone administration. To minimize participant burden and frustration, we established failure rules for each item and a stopping rule after failure of 5 consecutive items (with the remaining items scored as failed).

We documented whether each item was administered validly, not completed because of severe global impairment caused by the TBI, or not completed because of limitations unrelated to the TBI (eg, jaw wired shut). Missing scores because of global impairment were converted to failing scores for analysis.

FIM and DRS scores were obtained within 3 days of administration of the other items. For inpatients, they were abstracted from the medical record or provided by treating clinicians. For outpatients they were determined by structured interview of the participant or caregiver.

Data analysis

Responses for continuously scored items were assigned into 3–4 ordinal categories reflecting ascending levels of functioning. Where possible, categories were based on published norms. Where norms were not available, 2 of the authors (Y.B., M.S.) created categories based on visual inspection of the score distributions and affirmed by coauthors. Scoring categories were not intended to distinguish gradations within the normal range. Therefore, where norms were available, all scores within 1 SD of the mean were assigned to the highest category, with the remaining scores categorized into approximately equal-sized groups.

Item response theory was used in the analysis because this model is uniquely suited to our objective of combining items from multiple instruments with different response scales into an equal-interval, unidimensional interval measure. We used the Rasch partial credit model[26] such that items from the same source instrument with the same response categories were regarded as sharing the same rating scale. Rasch analysis provides estimates of internal consistency, item fit to the Rasch model, rating scale properties, and evaluation of dimensionality.[26] We evaluated rating scale performance using Fisher criteria, including internal consistency (criterion: person separation reliability ≥0.80), item fit to the Rasch model (criterion: mean square infit statistic <1.4), and unidimensionality (criterion: residual variance from the first principle components analysis <10%).[24] We used Winsteps software[a] to complete the rating scale analysis.[26,27] The available sample is sufficient to estimate item location within +½ logit with 99% confidence.[28]

Results

Table 2 presents a summary of the iterative Rasch analyses. We began by analyzing the original 18 FIM items with their original 7-category rating scale (see table 2, row 1) as a basis of comparison for the new measure. While person and item reliability were excellent

(0.92, 0.94), the rating scale was not monotonic in that several intermediate rating scale categories were never the most likely response across the range of measures. Further, 3 items misfit the Rasch model, 9.2% of the sample was at the floor of the measure and about 5% at the ceiling, and multiple dimensions were evident.

Row 2 reports analysis of all 49 administered items with their original rating scales. These items provided high reliability and minimal ceiling/floor issues, but many items had misordered rating scales. In subsequent analyses, we tried several strategies to maximize the psychometric properties of the item set, first by rescoring rating scales with misordered categories, then by deleting items.

Row 3 shows rescoring the FIM items into 3 categories rather than 7 improved FIM item fit to the model, but 2 items had misordered rating scales, many items misfit, and the residual analysis revealed a significant subfactor. Although the measure had less floor effect than the FIM, the ceiling effect was similar.

Row 4 shows deleting 18 misfitting items and rescoring the FIM and CRS-R items left 1 item with a misordered rating sale, no misfitting items, but 13% residual variance in the first principle component and a sizable ceiling effect (10.9%), greater than the FIM itself.

Row 5 shows that given the primary aim of the project to extend the range of the scale, we returned to the full set of 49 items in an attempt to preserve the improvements in floor and ceiling effects seen in row 2 while also improving the psychometric properties. Retaining all items and rescoring rating scales aggressively left 1 item with a misordered rating scale, 7 misfitting items, and a factor with residual variance exceeding the 10% threshold. The floor effects were satisfactory, but ceiling effects were again similar to the FIM.

Row 6 reports deleting 22 items and applying the rescoring from row 5 leaves 1 misordered rating scale and residual variance slightly above the 10% criterion but increases the ceiling effect.

In row 7 we decided to tolerate poor monotonicity in the FIM items' rating scales, in view of the prevalent use of the 7-point scale in practice, and returned to the 31 items included in row 4's analysis while rescoring the CRS-R items. Misordered rating scale categories were evident for the FIM items and 2 other items (CRS-R Motor Function Scale, DRS Motor Response). The ceiling effect was similar to the FIM with substantial improvement in the floor effect. No items misfit, and the residual variance only slightly exceeded the 10% criterion.

Row 8 shows with the rescoring of 2 additional items, both ceiling and floor effects were substantially improved with respect to the FIM, no items misfit, and the residual variance only slightly exceeded the 10% criterion. This set of items enhances rating scale monotonicity (except for the FIM, CRS-R Motor Function Scale, and DRS Motor Response items), minimizes multidimensionality, obviates item misfit, and minimizes floor and ceiling effects. Whereas a total of 14.1% of the sample was measured at either floor or ceiling with the FIM, this was reduced to 2.7% (all at ceiling) for the Brain Injury Functional Outcome Measure (BI-FOM). An item map that places these 31 items in order of difficulty

on the underlying logit scale and the conversion from the raw item score to the underlying Rasch score (transformed to a 100-point scale) are provided as supplemental figure S1 and supplemental table S3, respectively (available online only at http://www.archives-pmr.org/).

Discussion

The 31-item version of the BI-FOM substantially increases the range of measurement of functional ability after TBI compared with the FIM, surpasses its person and item reliability, and eliminates its floor effects while substantially reducing ceiling effects. Although the BI-FOM shows some evidence of an additional dimension, this is less prominent than for the FIM itself, which is widely used in research and clinical monitoring. The BI-FOM data fit the Rasch model relatively well, allowing its use as an interval level measure.

The combination of items targeting different ranges of function resulted in a measure with a logit range of 14.98 (vs 11.71 for FIM) and 2.7% of participants at ceiling (vs 4.9% at ceiling and 9.2% at floor for FIM). The remaining ceiling effects may reflect the previously discussed difficulty in finding normative items that tap the highest levels of function without introducing additional measurement dimensions. Indeed, a number of more difficult cognitive items were removed because of item misfit. The binning decisions made for the cognitive items may also have contributed to the residual ceiling effect. Scores from 1 SD below the mean extending to the highest possible score were all collapsed into a single category reflecting normal functioning. This is consistent with FIM scoring, which does not distinguish normal independence from exceptional physical and cognitive abilities. While it may be true that 2.7% of those with moderate-severe TBI recover to a "normal" level of global function as defined here, future research focused on improving sensitivity at the upper range could try binning that distinguishes superior from normal function and/or addition of more difficult items.

Study limitations

This study has several limitations. First, the TBIMS comprise specialized centers for inpatient rehabilitation, and thus, our findings may not generalize to individuals treated at less specialized centers or not receiving rehabilitation. However, the quality and intensity of treatment is more likely to affect movement along a recovery dimension than the structure of that recovery dimension itself. This measure was developed on a sample of individuals with moderate-severe injury, and although it captures extensive recovery in those individuals, we do not know its sensitivity to injury and recovery in patients with mild TBI. Moreover, despite attempts to enroll participants with a wide range of functional severity, patients with DOC are underrepresented in acute inpatient rehabilitation settings and were few in the study sample. Several BI-FOM items demonstrated evidence of nonmonotonicity, which may have been partly attributed to a small sample size in certain rating categories. Future studies with larger samples of patients with DOC should examine the need for further collapsing of these items. Moreover, several of the items assessing functioning in this subpopulation must be administered in person. Given that most studies will enroll and examine participants with DOC while they are still in an institutional setting and that by the time of later outcome assessment the vast majority will be able to complete telephone-based

measures, we believe this is an acceptable limitation. Moreover, research to validate a telephone version of the CRS-R is underway.

The BI-FOM also demonstrated evidence of slight multidimensionality. However, given that this was only slightly above the cutoff of 10% (and lower than with the FIM itself), that the item and person reliability were excellent, and that the intent of the scale is to assess a global functional dimension, we believe this level of multidimensionality is acceptable.

Because the BI-FOM items have been taken from a variety of instruments that have been translated into other languages and validated across cultures, the BI-FOM will likely also have value in international comparative research. This use in various languages raises the question of cross-cultural validity, which addresses how well the items on a translated or culturally adapted outcome measure reflect the performance of the items on the original instrument. Testing this assumption would require administration of the BI-FOM to different language groups (eg, English, Spanish), and then performing either confirmatory factor analysis or item response theory—based analyses (ie, invariant item ordering/differential item functioning) to assess equivalence. This is a worthwhile downstream aim but one that would require additional prospective data collection.

Conclusions

If replicated in an independent sample, the BI-FOM's ability to reliably measure global function across almost 15 logits will make it well suited as a single measure of outcome in longitudinal research and treatment trials intended to affect global functioning. The BI-FOM may also offer a clinically meaningful and useful tool for monitoring recovery and response to rehabilitation over time. Beyond research, the BI-FOM could inform clinical practice and care management, help determine candidacy for potential interventions, and better inform prognostic and care planning conversations with families affected by moderate-severe TBI. Other measures of functioning will continue to be useful for more precise assessment of a particular level of functioning (eg, the full CRS-R to assess patients with DOC) or of a more focused functional domain (eg, use of the full Brief Test of Adult Cognition by Telephone to assess cognitive function). While developed in the context of inpatient rehabilitation and 1-year follow-up, the instrument may also be appropriate for research on moderate-severe TBI in acute care settings and extending beyond 1 year in community or residential settings. The feasibility of administration is increased by the fact that all except the CRS-R items may be administered by phone, and research on telephone validation of the CRS-R is underway. The use of Rasch analysis provides an interval level measure appropriate for applying advanced statistical analysis techniques.

Supplier

a. Winsteps, Winsteps Rasch Measurement Computer Program.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript

Acknowledgments

List of abbreviations:

BI-FOM	Brain Injury Functional Outcome Measure
CRS-R	Coma Recovery Scale—Revised
DOC	disorders of consciousness
DRS	Disability Rating Scale
TBI	traumatic brain injury
TBIMS	Traumatic Brain Injury Model System

References

1. Whyte J, Ponsford J, Watanabe T, et al. In: Frontera WR, editor. Traumatic brain injury In: DeLisa's physical medicine and rehabilitation: principles and practice. 5th ed. Philadelphia: Lippincott Williams & Wilkins Health; 2010. p 575–623.

2. Giacino JT, Katz DI, Schiff ND, et al. Practice guideline update recommendations summary: disorders of consciousness. Arch Phys Med Rehabil 2018;99:1699–709. [PubMed: 30098791]

3. Nakase-Richardson R, Whyte J, Giacino J, et al. Longitudinal outcome of patients with disordered consciousness in the NIDRR TBI Model Systems programs. J Neurotrauma 2012;28:1–8.

4. Wilde EA, Whiteneck GG, Bogner J, et al. Recommendations for the use of common outcome measures in traumatic brain injury. Arch Phys Med Rehabil 2010;91:1650–60. [PubMed: 21044708]

5. Rappaport M, Hall KM, Hopkins K, et al. Disability rating scale for severe head trauma: coma to community. Arch Phys Med Rehabil 1982;63:118–23. [PubMed: 7073452]

6. Hall K, Hamilton B, Gordon W, et al. Characteristics and comparisons of functional assessment indices: Disability Rating Scale, Functional Independence Measure, and Functional Assessment Measure. J Head Trauma Rehabil 1993;8:60–74.

7. Granger CV, Divan N, Fiedler RC. Functional assessment scales: a study of persons after traumatic brain injury. Am J Phys Rehabil 1995; 74:107–13.

8. Hall KM, Bushnik T, Lakisic-Kazazic B, et al. Assessing traumatic brain injury outcome measures for long-term follow-up of community-based individuals. Arch Phys Med Rehabil 2001;82:367–74. [PubMed: 11245760]

9. Hammond FM, Grattan KD, Sasser H, et al. Long-term recovery course after traumatic brain injury: a comparison of the functional independence measure and disability rating scale. J Head Trauma Rehabil 2001;16:318–29. [PubMed: 11461655]

10. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. Arch Phys Med Rehabil 2004;85:2020–9. [PubMed: 15605342]

11. Wilson J, Pettigrew L, Teasdale G. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: guidelines for their use. J Neurotrauma 1998;15:573–85. [PubMed: 9726257]

12. Bagiella E, Novack T, Ansel B, et al. Measuring outcome in traumatic brain injury treatment trials: recommendations from the TBI Clinical Trials Network. J Head Trauma Rehabil 2010;25:375–82. [PubMed: 20216459]

13. Malec JF. Rehabilitation research: 30 years later; 30 years hence. J Head Trauma Rehabil 2013;28:227–31. [PubMed: 23661076]

14. Lachman ME, Agrigoroaei S, Tun PA, et al. Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. Assessment 2014;21:404–17. [PubMed: 24322011]

15. Giacino JT, Whyte J, Bagiella E, et al. Placebo-controlled trial of amanatadine for severe traumatic brain injury. N Engl J Med 2012; 366:819–26. [PubMed: 22375973]

16. Sherer M, Nakase-Thompson R, Yablon SA, et al. Multidimensional assessment of acute confusion after traumatic brain injury. Arch Phys Med Rehabil 2005;86:896–904. [PubMed: 15895334]

17. Bodien YG, Martens G, Ostrow J, et al. Cognitive impairment, clinical symptoms and functional disability in patients emerging from the minimally conscious state. NeuroRehabilitation 2020;46:65–74. [PubMed: 32039868]

18. Hart RP, Levenson JL, Sessler CN, et al. Validation of a cognitive test for delirium in medical ICU patients. Psychosomatics 1996;37:533–46. [PubMed: 8942204]

19. Kennedy R, Nakase-Thompson R, Sherer M, et al. Use of the cognitive test for delirium in patients with traumatic brain injury. Psychosomatics 2003;44:1–7. [PubMed: 12515832]

20. Stuss DT, Binns MA, Carruth FG, et al. The acute period of recovery from traumatic brain injury: posttraumatic amnesia or posttraumatic confusional state? J Neurosurg 1999;90:635–43. [PubMed: 10193606]

21. Nakase-Thompson R, Manning E, Sherer M, et al. Brief assessment of severe language impairments: initial validation of the Mississippi Aphasia Screening Test. Brain Inj 2005;19:685–91. [PubMed: 16195182]

22. Boake C Supervision Rating Scale: a measure of functional outcome from brain injury. Arch Phys Med Rehabil 1996;77:765–72. [PubMed: 8702369]

23. Whiteneck G, Dijkers M, Heinemann AW, et al. Development of the Participation Assessment with Recombined Tools-Objective for use with individuals with traumatic brain injury. Arch Phys Med Rehabil 2011;92:542–51. [PubMed: 21367393]

24. Gordon W, Mann N, Willer B. Demographic and social characteristics of the traumatic brain injury model system database. J Head Trauma Rehabil 1993;8:26–33.

25. Thibaut A, Bodien YG, Laureys S, et al. Minimally conscious state "plus:" diagnostic criteria and relation to functional recovery. J Neurol 2020;267:1245–54. [PubMed: 31773246]

26. Bond T, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. London: Routledge; 2015.

27. Fisher WP. Rating scale instrument quality criteria. Rasch Meas Trans 2007;21:1095.

28. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994;194(7):328.
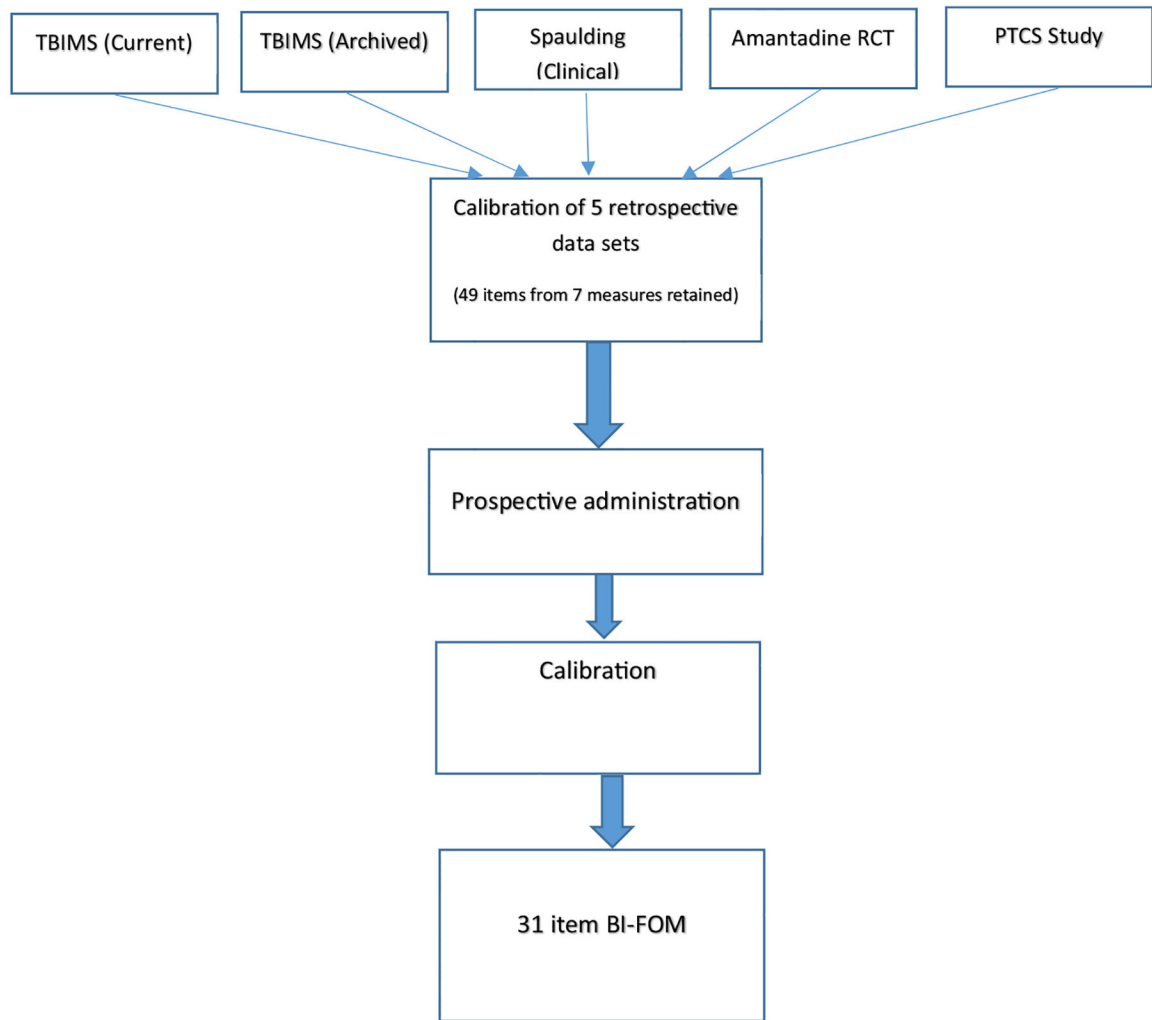
Fig 1.
Sequence of steps in this multistep study, beginning with retrospective analysis of 5 existing data sets and proceeding on to prospective administration and calibration of assessment items. Abbreviations: PTCS, posttraumatic confusional state; RCT, randomized controlled trial.
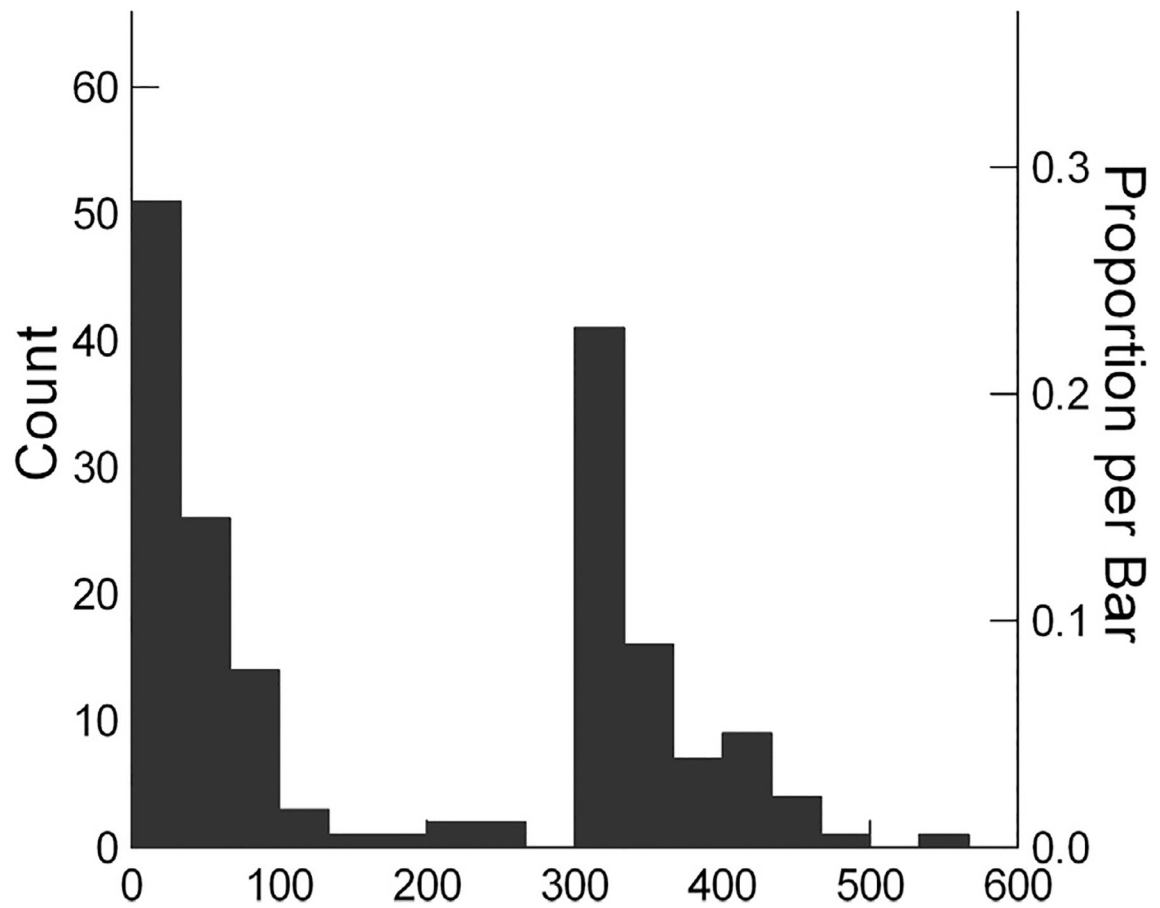
Fig 2.
Range of times post injury when assessment data were collected. The x-axis represents time post injury and the y-axis the number and proportion of the prospective sample who were assessed at each time point.

Table 1

Participant characteristics

Characteristics	n	Value
Age (y), mean, median (range)	169	42.57, 42.0 (16–89)
Cause of injury, n (%)	169	
- Vehicular+pedestrian		97 (57.4)
- Fall		50 (29.6)
- Gunshot		11 (6.5)
- Other		11 (6.5)
Time in PTA (d), mean, median (range)	149	18.22, 13.0 (0–79)
Acute hospital LOS (d), mean, median (range)	169	17.64, 15.0 (2–70)
Rehab hospital LOS (d), mean, median (range)	169	30.35, 21.0 (4–202)
FIM at assessment, mean, median (range)	184	87.93, 102.5 (18–126)
Time from injury to assessment (d), mean, median (range)	179	182.19, 95.0 (4–557)

Abbreviations: LOS, length of stay; PTA, posttraumatic amnesia.

Table 2

Summary of Rasch partial credit analyses

Analysis	Items, Categories, People	Person, Item Reliability	% of Sample at Ceiling & Floor Max & Min Logit Value	Misordered Rating Scales	Misfitting Items With Mean Square Infit >1.40	Item-Measure r<.4	Factor 1 Variance >10%	Notes
1. FIM items, no rescoring	18i 126c 185p	0.92 0.94	C 4.9% F 9.2% Max 6.80 Min −4.91	Feeding Bladder Management Bed/Chair/ Wheelchair Transfers Stairs Social Interaction	Memory Walk/Wheelchair Expression	None	21.1%	
2. All items with original scoring	49i 244c 184p	.97 .99	C 0% F 0.5% Max 7.47 Min −6.59	Visual Function Scale CRS-R Motor Function Scale Conceptual Reasoning 18 FIM Items DRS Motor Response	Short Delay Word Recall Auditory Vigilance (simple) Auditory Vigilance (complex) Vigilance Subtest A Auditory Function Scale Visual Function Scale Arousal Scale	Auditory Vigilance (complex)	11.4%	While reliability is high and ceiling/floor issues are minor, there are multiple rating scale and item misfit issues.
3. All items, FIM items rescored	49i 173c 184p	0.97 0.99	C 4.3% F 0.5% Max 6.19 Min −7.36	Visual Function Scale CRS-R Motor Function Scale	Short Delay Word Recall Auditory Vigilance (simple) Auditory Vigilance (complex) Vigilance Subtest A Auditory Function Scale Visual Function Scale Arousal Scale	None	15.9%	Rescoring FIM items: 1, 2–5, 6–7 eliminates FIM rating scale monotonicity problems without resolving item misfit.
4. 18 items deleted, FIM and CRS items rescored FIM+CRS Visual, Motor rescored+Delete Mis-OverFit3.out	31i 103c 184p	0.95 0.99	C 10.9% F 0.5% Max 7.50 Min −9.37	CRS-R Motor Function Scale	None	None	13.0%	Visual Function Scale rescored: 0, 1, 2, 3, 4–5 CRS-R Motor Function Scale rescored: 0–1, 2, 3–5, 6 FIM Items rescored: 1, 2–5, 6–7 Deleted: 9 misfitting & 9 overfitting items from previous analysis 9 misfits: Word List Recall (short delay), Auditory Vigilance (simple & complex), Vigilance Subtest

A, Visual Function Scale, Auditory Function Scale, Arousal Scale, Controlled Oral Word Association Test (FAS), Backward Counting (from 100)

9 overfit: Toileting, Problem Solving, Bed/Chair/Wheelchair Transfers, Dressing Upper Body, Conceptual Reasoning, DRS Grooming, DRS Feeding, DRS Toileting, FIM Toilet Transfers

Analysis	Items, Categories, People	Person, Item Reliability	% of Sample at Ceiling & Floor Max & Min Logit Value	Misordered Rating Scales	Misfitting Items With Mean Square Infit >1.40	Item-Measure $r<.4$	Factor 1 Variance >10%	Notes
5. All items, various item rescoring	49i 144c 184p	0.96 0.98	C 4.3% F 1.1% C 7.44 F −9.33	DRS Motor Response	Vigilance Subtest A; Backward Counting (from 100); Auditory Vigilance (simple & complex); Visual Function Scale; Auditory Function Scale; CRS-R Motor Function Scale	None	14.5%	FIM Items rescored. Rescored for this and all subsequent analyses: Vigilance Subtest A: 0–35, 36; Auditory Vigilance (simple): 0, 1–18, 19–20; Auditory Vigilance (complex): 0, 1–8, 9–10; Count to 40 by 3s: 0, 1–11, 12–13; Word List Recall (short delay): 0, 1–2, 3+; Rescored for this analysis only: Eye Opening 0: None/To Pain/To Speech; 1: Spontaneous; Communication Ability 0: None/Incomprehensible; 1: Inappropriate/Confused; 2: Oriented; DRS Toileting 0: None 1: Minimal/Partial, 2: Complete
6. 22 items deleted	27i 81c 184p	0.94 0.98	C 12.5% F 4.3% Max 6.99 Min −10.54	DRS Motor Response	None	None	12.9%	Deleted: 1–6, 15–18, 20, 21, 23, 27–29, 32, 33, 42, 43, 46, 47 (FIM Items rescored)
7. 18 items deleted, CRS item rescored	31i 155c 184p	0.96 0.99	C 4.9% F 0.5% Max 7.24 Min −8.28	CRS-R Motor Function Scale; DRS Motor Response; 13 FIM Items	None	None	10.8%	Visual Function Scale Rescored: 0, 1, 2, 3, 4; CRS-R Motor Function Scale Rescored: 0–1, 2, 3–5, 6; FIM Items: 7 point scale (no rescoring); Deleted: 9 misfitting & 9 overfitting items from previous analyses; 9 misfits: Word List Recall (short delay), Auditory Vigilance (simple & complex), Vigilance Subtest A, Visual Function Scale, Auditory Function Scale, Arousal Scale, Controlled Oral Word Association Test (FAS), Backward Counting (from 100); 9 overfit: FIM Toileting, Problem Solving, Bed/Chair/Wheelchair Transfers, Dressing Upper Body, Conceptual Reasoning, DRS Grooming, DRS Feeding, DRS Toileting, Toilet Transfers
8. 18 items deleted, CRS, 2 additional items rescored	31i 153c 185p	0.96 0.99	C 2.7% F 0% Max 7.26 Min −7.72	CRS-R Motor Function Scale; 4 FIM Items: Feeding, Bladder Management,	None	None	10.9%	18 Items deleted: listed in #7 above. CRS Item rating scales rescored as above, plus Auditory: Total correct binned 0–3, 4; Count to 40 by 3s 0, 1–11, 12–13

Analysis	Items, Categories, People	Person, Item Reliability	% of Sample at Ceiling & Floor Max & Min Logit Value	Misordered Rating Scales	Misfitting Items With Mean Square Infit >1.40	Item-Measure r<.4	Factor 1 Variance >10%	Notes
				Stairs, Social Interaction				

NOTE. Column 1 identifies the analysis; column 2 identifies the number of items, rating scale categories, and participants in the analysis; column 3 reports the person and item reliability; column 4 reports the percentage of the sample at the ceiling and floor of the measure as well as the maximum and minimum observed logit values; column 5 lists items with misordered rating scales; column 6 lists misfitting items; column 7 lists items correlating <0.4 with the measure; column 8 reports the percentage of variance accounted for by the first principal component analysis of residual information; and column 9 provides notes about item rescoring and deletion decisions.

Abbreviations: c, category; C, ceiling; F, floor; i, item; p, people.