Journal of Neurotrauma
38:2419-2434 (September 1, 2021)
Mary Ann Liebert, Inc.
DOI: 10.1089/neu.2020.7528

Journal of Neurotrauma

ORIGINAL ARTICLE CLINICAL STUDIES

Central Curation of Glasgow Outcome Scale-Extended Data: Lessons Learned from TRACK-TBI

Kim Boase,[1,*] Joan Machamer,[1] Nancy R. Temkin,[1] Sureyya Dikmen,[1] Lindsay Wilson,[2] Lindsay D. Nelson,[3] Jason Barber,[1] Yelena G. Bodien,[4] Joseph T. Giacino,[4] Amy J. Markowitz,[5] Michael A. McCrea,[3] Gabriela Satris,[5] Murray B. Stein,[6] Sabrina R. Taylor,[5] and Geoffrey T. Manley[5]; and the TRACK-TBI Investigators**

Abstract

The Glasgow Outcome Scale (GOS) in its original or extended (GOSE) form is the most widely used assessment of global disability in traumatic brain injury (TBI) research. Several publications have reported concerns about assessor scoring inconsistencies, but without documentation of contributing factors. We reviewed 6801 GOSE assessments collected longitudinally, across 18 sites in the 5-year, observational Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study. We recorded error rates (i.e., corrections to a section or an overall rating) based on site assessor documentation and categorized scoring issues, which then informed further training. In cohort 1 ($n = 1261$; February 2014 to May 2016), 24% of GOSEs had errors identified by central review. In cohort 2 ($n = 1130$; June 2016 to July 2018), acquired after curation of cohort 1 data, feedback, and further training of site assessors, the error rate was reduced to 10%. GOSE sections associated with the most frequent interpretation and scoring difficulties included whether current functioning represented a change from pre-injury (466 corrected ratings in cohort 1; 62 in cohort 2), defining dependency in the home and community (163 corrections in cohort 1; three in cohort 2) and return to work/school (72 corrections in cohort 1; 35 in cohort 2). These results highlight the importance of central review in improving consistency across sites and over time. Establishing clear scoring criteria, coupled with ongoing guidance and feedback to data collectors, is essential to avoid scoring errors and resultant misclassification, which carry potential to result in "failure" of clinical trials that rely on the GOSE as their primary outcome measure.

Keywords: central review; clinical outcome assessments; data curation; GOSE; traumatic brain injury

Introduction

The Glasgow Outcome Scale, in either its original (GOS)[1] or extended form (GOSE),[2] is the most widely used outcome measure in traumatic brain injury (TBI) research today, with >4000 citations to the original article describing the GOS.[3] The GOSE is a core National Institute of Neurological Disorders and Stroke (NINDS) TBI Common Data Element,[4,5] indicating that it is recommended in all types of TBI research involving adults, including observational studies and clinical trials sponsored by NINDS. It has also been accepted by the U.S. Food and Drug Administration as the primary end-point of efficacy for TBI drug trials.

[1]Department of Neurological Surgery, Harborview Medical Center, University of Washington, Seattle, Washington, USA.
[2]Division of Psychology, School of Natural Sciences, University of Stirling, Stirling, United Kingdom.
[3]Department of Neurological Surgery, Medical College of Wisconsin, Milwaukee, Wisconsin, USA.
[4]Spaulding Rehabilitation Hospital Massachusetts General Hospital, Charlestown, Massachusetts, USA.
[5]Brain and Spinal Injury Center, University of California, San Francisco, San Francisco, California, USA.
[6]Department of Psychiatry, University of California, San Diego, La Jolla, California, USA.
**The TRACK-TBI Investigators may be found at the end of this article.

*Address correspondence to: *Kim Boase, BA, Department of Neurological Surgery, Harborview Medical Center, Box 359924, 325 9th Avenue, Seattle, WA 98104, USA E-mail:* kboase@uw.edu

© Kim Boase *et al.*, 2021; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons Attribution Noncommercial License [CC-BY-NC] (http://creativecommons.org/licenses/by-nc/4.0/) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

For many years, concerns have been raised about the inter-rater reliability of both the GOS and the GOSE, which may vary depending on several factors, including the type of assessor (e.g., primary care physician, intensive care unit [ICU] physician, or psychologist) and their proficiency in administering the measure.[6–10] In 1998, a semistructured interview was developed to provide the assessor with initial and follow-up questions for the scales,[2] as a way to reduce inter-rater variability and improve sensitivity. Although the GOSE interview helped improve inter-rater reliability, inconsistencies have remained an issue,[11–13] with inter-rater variation ranging from 17%[13] to 40%.[11] Further refinements to the structured interview, and a guide for the administration of the GOSE, drawn from assessor experience in the Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) and Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) studies, now appear in a *Manual for the Glasgow Outcome Scale-Extended (GOSE) Interview*, developed by Wilson and colleagues.[14]

Use of coarse outcome measurement has been posited as one explanation for the persistent failure of TBI pharmaceutical trials to identify beneficial treatments.[15,16] The GOSE, a primary outcome end-point, can be administered in different ways, including whether the assessor includes the effects of polytrauma or TBI only, which, in turn, can lead to inconsistency in assigning the overall GOSE rating. The approach to administering outcome assessments, the assessor's training, and adherence to protocol guidelines can vary across centers and studies. These variations point to concerns with reliability and accuracy in multi-site studies, leading to inconsistent outcome results.[8] Further, methods of collecting outcomes are often inadequately documented in published studies.[17] The recently published Guidelines for Data Acquisition, Quality and Curation for Observational Research Designs moved to fill this gap.[18] Though the importance seems clear, little had been published specifically on steps to facilitate or ensure high data quality in collection of TBI outcome measurement and, particularly, the GOSE.

Initial training is paramount, but is not entirely sufficient. One way to improve assessor accuracy in scoring of the GOSE is through central review of assessments. Wilson and colleagues[11] found a marked decline in data queries after an initial period of review, feedback, and training in the multi-center efficacy trial of dexanabinol on TBI outcome. Lu and colleagues[19] also recommended central curation as a way to reduce error rates on the GOSE. However, the details of the typical errors and inconsistencies that were uncovered in these studies were not reported.

We conducted a comprehensive review of GOSE assessments in the multi-site, longitudinal, observational Transforming Research and Clinical Knowledge in TBI study (TRACK-TBI).[20] The independent central review examined error rates, types of errors, and extent of change (the number of GOSE points change from the original rating to the curated rating) in GOSE ratings in two time periods of the study. The aim of this work was to improve the accuracy and consistency of post-TBI functional status ratings among GOSE data collectors. A secondary goal was to identify the assessment areas within the GOSE that caused significant variation in scoring and could lead to misclassification of functional outcome. Until now, many investigators relied on the 1998 journal article introducing the semistructured interview format as guidance for the administration of the GOSE. Variation in interpretation of particular sections is likely one explanation for the difficulties encountered in the administration and overall inconsistency of the measure. The publication of the *Manual for the Glasgow Outcome Scale-Extended (GOSE) Interview* will bring further clarity to the field.[14]

Identification of these difficult areas helped to create training protocols to improve consistency of administration and scoring of this primary outcome measure. TBI investigators responsible for developing training materials and reviewing GOSE data may benefit from our experience and from the newly published GOSE administration and scoring manual, which also appears in this issue of the *Journal of Neurotrauma*.[14]

Methods

Participants

TRACK-TBI is a prospective, observational study that enrolled 2698 TBI patients across the life span (ages 0–99 years) and spectrum of injury from severe to mild (Glasgow Coma Scale [GCS] score 3–15), from February 26, 2014 through July 27, 2018. English- or Spanish-speaking participants were enrolled within 24 h of injury at 18 U.S. Level I trauma centers. All participants had a head computed tomography scan ordered as part of their clinical care. Full inclusion and exclusion criteria can be found on the TRACK-TBI Web page.[20] Participants were followed at 2 weeks and 3, 6, and 12 months post-injury. Surviving participants ≥17 years of age with TBI were included in this analysis if at least one of the four GOSE follow-up assessments was completed and the participant had not withdrawn consent. We divided the analysis into two time periods. From February 2014 to May 2016, TBI cases were recruited from the first 11 U.S. Level 1 trauma centers that participated in the study. These participants comprise cohort 1 ($n = 1261$). Beginning in June 2016, seven more centers were added to the study and participants were enrolled across the 18 sites between then and July 27, 2018 (cohort 2; $n = 1130$).

Following the study objectives, cohort 1 enrolled TBI participants across three care pathways: roughly one third who were discharged directly from the emergency

department (ED); one third who were admitted to the hospital but not the ICU; and one third who were admitted to the ICU. This distribution of cases resulted in a sample with mostly milder brain injuries. For cohort 2, the study objectives shifted to preferentially enroll those with more-severe injuries. All sites continued enrollment of patients admitted to the ICU or hospital. Only 7 of the 18 sites which were participating in a magnetic resonance imaging substudy continued enrollment of participants discharged from the ED. The study was approved by the institutional review board at each site, and all participants either consented for themselves or were consented by a legally authorized representative.

Assessments

Glasgow Outcome Scale-Extended. The GOSE is an 8-point scale representing levels of functioning ranging from death (1) to upper good recovery (8). The assessment is based on change from the pre-injury level of functioning and is administered using a semistructured interview format.[2] The GOSE interview consists of standard questions covering eight areas of function; however, the interviewer is expected to ask additional questions to glean the information required to assess limitations within a specific section. The eight sections are: 1) level of consciousness; 2) assistance within the home; 3) independence outside the home for shopping; 4) independence outside the home for travel; 5) return to work or school; 6) social and leisure participation; 7) close relationships; and 8) return to normal life, as relates to symptoms. The assessor uses the responses to assign a functional level rating of vegetative state (rating of 2), lower or upper severe disability (rating of 3 or 4), lower or upper moderate disability (rating of 5 or 6), or lower or upper good recovery (rating of 7 or 8).

The overall GOSE rating is determined according to the lowest (worst) rating assigned in any section for which the response signifies a change in function from pre-injury status. All sections of the GOSE were administered, irrespective of answers provided on earlier sections, unless the participant was deceased or in a vegetative state. Participants with disturbance in consciousness were evaluated using the Coma Recovery Scaled-Revised[21] to determine whether they were in a vegetative state.

As discussed by Wilson and colleagues (1998),[2] the GOSE can be administered either to include effects of peripheral injuries and other consequences of the injury in addition to the effects of the TBI (referred to here as GOSE-All) or by parsing the effects of the TBI only (GOSE-TBI).[22] In either case, the rating reflects change from pre-injury status. Initially, TRACK-TBI considered only the effects of the TBI in the GOSE ratings. Approximately 9 months into data collection, the protocol was changed and both ratings—GOSE-All and GOSE-TBI—were obtained at each assessment. In each section, except for that assessing level of consciousness, the assessor first queried the level of functional limitation associated with the overall injury and then asked the interviewee to assess the functional limitations attributed to just the TBI. This was referred to by the TRACK-TBI study, and hereafter in this article, as the ''GOSE 2-ways interview.'' The lowest (worst) rating in any section that indicated a change from pre-injury function determined the overall GOSE-All rating. The overall GOSE–TBI rating was based on the lowest (worst) rating in any section attributed to only the TBI. The decision to obtain the two GOSE scores was made to allow comparison of the TRACK-TBI results with those from the longitudinal, observational CENTER-TBI study,[23] which acquired only GOSE-All ratings.

Patient/Surrogate interviews. An interview completed at the time of enrollment, by the participant or their surrogate, collected pre-injury information, including work and school status and living situation.

At each follow-up period, participants, or a close family member in the case of the more severely injured, completed a post-injury follow-up interview. This interview documented current status and the reason for any change from the pre-injury situation. Both interviews were created by the TRACK-TBI outcome core leadership team guided by NINDS TBI Common Data Elements.

Assessor training

Before opening enrollment, outcome assessors from the first 11 sites received instruction in the administration of the GOSE and other outcome measures at a 2-day in-person training session convened by TRACK-TBI's outcome core leadership and conducted by appointed team members. Nine months into data collection, training was conducted by telephone to introduce the GOSE 2-ways assessment. When the seven new sites were added, another in-person training session was held for assessors from all 18 sites. The second session was designed to address problem areas in administering and scoring the GOSE that had been identified by the TRACK-TBI Outcomes Core leadership team. The Standard Operating Procedures for Outcome Assessment Manual (SOP),[24] with specific guidance for the administration of the GOSE based on material codified in the Wilson and colleagues Manual published concurrently,[14] was posted to the TRACK-TBI Web site and provided to assessors.

All assessors joining the study at any time were required to practice administering the GOSE with mock participants. TRACK-TBI outcome assessors ranged in level of education from a bachelor's degree to PhD or MD, with a BA or MA being the most common. Many outcome assessors had previous outcome evaluation experience as well as experience in the field of TBI. Assessors submitted a video demonstration of their mock administration, which was reviewed by a central review team member (S.T., K.B.) who confirmed their

competence to administer the GOSE. Thereafter, source documents (paper copies) of two cases from each assessor were reviewed, and the data, as entered into the electronic database, were double checked. Assessors were instructed to record significant information influencing the determination of a section rating on the source document. In addition to this, a text field was added to the GOSE in the data capture system, allowing this pertinent information to be available for review and remain as part of the permanent record.

Random checks of ~10% of all GOSEs administered per site were conducted by reviewing the database for logical consistency among items on the GOSE and comparing GOSE responses with the pre-injury interview and the post-injury follow-up interview(s). For example, a participant coded as currently living alone on the Patient/Surrogate interview but coded as dependent within the home on the GOSE would result in a query. The results of this database review, intended as another training opportunity, were sent to site assessors, noting any discrepancies, scoring inconsistencies, missing data, or data entry errors. Site assessors were asked to review their data, resolve the inconsistencies or errors, and make any necessary corrections. Monthly conference calls for site outcome assessors with the TRACK-TBI Outcomes Core leadership team provided continued training, with discussion of ambiguous situations and cases as well as review of training scenarios and SOPs.

Curation procedure

Figure 1 shows the curation process used to review GOSEs.

Curation of all GOSEs began ~2 years into the study. An audit log was created within the electronic database. Beginning with the first participant and working through the data set, each GOSE assessment for every participant was reviewed. Central reviewers (K.B., J.M., and G.S.) examined the pre-injury and post-injury follow-up interviews and the GOSE for consistency in the information recorded (Fig. 1). Any apparent inconsistencies, missing data points, or unusual combinations of responses were documented within the audit log. If there were no issues, the audit log was closed. Otherwise, queries were sent to the appropriate site for assessor review. The site assessor was asked to review source documents (paper copies) for relevant notes and confirm that data entry was correct. The assessor indicated within the audit log what action had been taken after this review. The options included: 1) correction made; 2) no changes needed, correct as is; or 3) no changes, insufficient information on the source documents to make a correction.

Two types of corrections could be made to the GOSE. The first was a correction to the rating in a specific section of the GOSE, for example, reduced work capacity (score = 6) changed to unable to work (score = 5). Such a change did not necessarily affect the overall rating, given that the overall rating may have been determined by another section within the GOSE that received a lower score. Changes to ratings in a specific section were tallied because the section ratings themselves have been used as outcomes.[25] The second type of correction was to the overall GOSE rating(s) (GOSE-All, GOSE-TBI, or both). The final determination as to whether a correction was warranted rested with the site assessors, not the central reviewers. Review of GOSEs ranged from 6 months to 3 years after administration for cohort 1. Cohort 2 reviews occurred from days to a few months after administration and completion of data entry.

Because the review process was not conducted immediately after the assessment, corrections to rating changes requiring judgment or subjective information relied on notes written contemporaneously with the evaluation by the assessor. In the case of those without documentation to justify a change in rating, the GOSE was not changed. It is important to note that all articles published by TRACK-TBI using GOSE data were submitted after the vast majority of the curation process had been completed.

Common scenarios in the query process to confirm consistency in data collection across the GOSE and the interviews are presented in Boxes 1 and 2 and Supplementary Box S1.

Box 1. Correction to the Overall GOSE Rating

Step 1. Central reviewer examines pre-injury interview for the following:
- **Pre-injury work/school status**: participant was employed prior to the injury
- **Pre-injury living situation**: living independently with spouse

Step 2. Central reviewer checks GOSE scores and free-text notes as entered into database by site assessor:
- **Independence inside and outside the home (Questions 2a, 3a, 4a)**: Participant rated as independent within the home, and with shopping and travel
- **Work (Question 5b)**: Currently working part-time (section score = 6)
- **Social and Leisure (Question 6b)**: Currently participating a bit less (section score = 7)
- **Relationships (Question 7b)**: Experiencing difficulties on a daily basis (section scored = 5)
  - Were there relationship difficulties before the injury? (Question 7c) Participant answered 'yes', *(This indicates the relationship problems existed prior to the injury and are not worse as a result of the injury; thus, no change due to the injury. Therefore, this section would not be used in assigning the overall rating.)*
- **Return to normal life/Symptom burden (Question 8a)**: endorses symptoms (section score = 7)

Overall rating assigned by assessor = 5 (lowest rating for any section)

Step 3. Central reviewer sends query to site assessor, as follows:
The coding of the relationship section indicates the difficulties precede the injury and are the same as before the injury (Question 7c). Therefore, the relationship section would not determine the overall rating. As the document is coded the overall rating should be 6, per the work section.

Step 4. Site assessor reviews paper documents for accuracy of data entry. Confirms and corrects the overall rating from 5 to 6, and indicates that a correction was made within the audit log. The correction to overall GOSE score is documented in the participant's record in the electronic data capture system.

Step 5. Central reviewer confirms the correction. Initials, dates, and closes audit log.

Box 2. Correction to a GOSE Section Resulting in no Change to Overall Rating.

(Same scoring scenario as Box 1, except that there is a free-text note by site assessor that appears inconsistent with the rating on the Relationship section)

- **Relationships** (Question 7b): Experiencing difficulties on a daily basis (section score = 5)

  Were there difficulties before the injury (Question 7c)? Coded 'yes', indicating that the relationship difficulties were present before the injury and have not worsened as a result of the injury.

*Free-text note reads, 'participant experienced difficulties with relationships before the injury but they are much worse now'.

Overall rating assigned by assessor = 5

Step 1. Central reviewer sends the following query to the site assessor:

> *The note entered into the database states that the difficulties with relationships are worse now. If that is the case, Question 7c (difficulties with relationships pre-injury) should have been coded ''no,'' to indicate that in this assessment the worsening of relationship difficulties did represent a change from pre-injury status. The overall rating would remain a 5 in this case. Please review.*

Step 2. Site assessor reviews documents and confirms that the relationship problems are worse now. Changes Q7c to ''no'' and notes this correction in the audit log. Correction is documented in the participant's record in the electronic data capture system.

Step 3. Central reviewer confirms the correction made. Initials, dates, and closes the audit log.

Statistical analysis

The analyses were largely descriptive. Weighted and unweighted kappa statistics were calculated to evaluate the degree of agreement between the original and post-curation scores. These analyses were performed using SAS software (version 9.4; SAS Institute Inc., Cary NC)

Results

Figure 2 presents the participant flow diagram of GOSE assessments for cohorts 1 and 2 across the study sites and time points. During the GOSE central review process, one site was identified as having systematically diverted from the GOSE administration protocol in significant ways, including the use of an alternate interview developed at their site, or making significant changes without the required documentation. For example, the assessor recorded peripheral injuries gleaned from the medical record after the completion of the assessment, and attributed functional changes to sections of the GOSE based on those injuries, despite no record of those functional limitations being reported by the participant at the time of the GOSE administration. After discussions with the site assessor and an in-person site visit, the TRACK-TBI Executive Committee decided to remove all GOSE data from that site from the central database ($n = 449$ GOSE assessments) until a new assessor was trained and certified. These removed GOSE assessments were not included in this analysis.

Table 1 presents characteristics of cohort 1 and cohort 2 participants with reviewed GOSEs. Injury severity covered the full range of the GCS, with a higher percentage of participants with low GCS in cohort 2, consistent with the shifting enrollment priorities of the study at this stage.

Frequency and magnitude of corrections to Glasgow Outcome Scale-Extended overall score and section ratings

The study's central reviewers examined 3668 GOSEs from 10 sites in cohort 1 and 3133 GOSEs from 18 sites in cohort 2 (Supplementary Table S1). In cohort 1, 1307 (36%) of interviews received a query and 867 (24%) resulted in a correction: 478 (13%) resulted in a change to one or both overall GOSE scores, 671 (18%) resulted in a change to a section rating, and 282 (8%) to both an overall rating and a section rating. In cohort 2, 625 (20%) GOSEs received a query, with 314 (10%) requiring a correction: 218 (7%) resulted in a change in one or both overall GOSE scores, 149 (5%) resulted in a correction to a section rating, and 53 (2%) resulted in both a section and overall rating change (cohort 2; Supplementary Table S1). Unlike cohort 1, reviews were done soon after data entry and substantial feedback had already been provided to site outcome assessors.

Queries and corrections also dropped consistently from early evaluations to the later ones. In cohort 1, 34% of the 2-week and 13% of the 12-month GOSEs required correction to an overall and/or section rating (Fig. 3). In cohort 2, 15% of the 2-week and 5% of the 12-month GOSEs required correction to an overall and/or section rating. Examining the overall rating only showed the same trend—the frequency of corrections in cohort 1 declined from 20% at 2 weeks to 7% at 12 months, and from 11% at 2 weeks to 3% at 12 months in cohort 2 (Fig. 4).

We also looked at the degree of change of the overall ratings (Table 2). In cohort 1, at 2 weeks, 172 participants were originally assigned an overall GOSE-All rating of 3 or 4 (severe disability). In 34 of these cases (20%), it was determined that the participant did not meet the criteria for dependency in the home, dependency with shopping or travel, or there was some other data error achieving an overall rating of 5 or better. The changes were, in most cases, attributable to orthopedic casts and whether the impact they have on functioning rises to the level of dependence specified in the SOP. By 6 months, the number had dropped to 55 participants receiving an overall GOSE-All rating of 3 or 4 and only 6 (11%) increasing to a rating of ≥5. In cohort 2, at 2 weeks, 257 participants initially received a rating of 3 or 4 with only 6 (2%) corrected to a rating of 5 or better. At 6 months, in cohort 2, 68 participants achieved an initial rating of 3 or 4 with only 1 (1%) corrected to a rating of 5 or better. Although the early time period rates of queries and changes in cohort 1 were high, calculating reliability according to kappa statistics, the lowest kappa was 0.83; 81% of unweighted kappas and all of

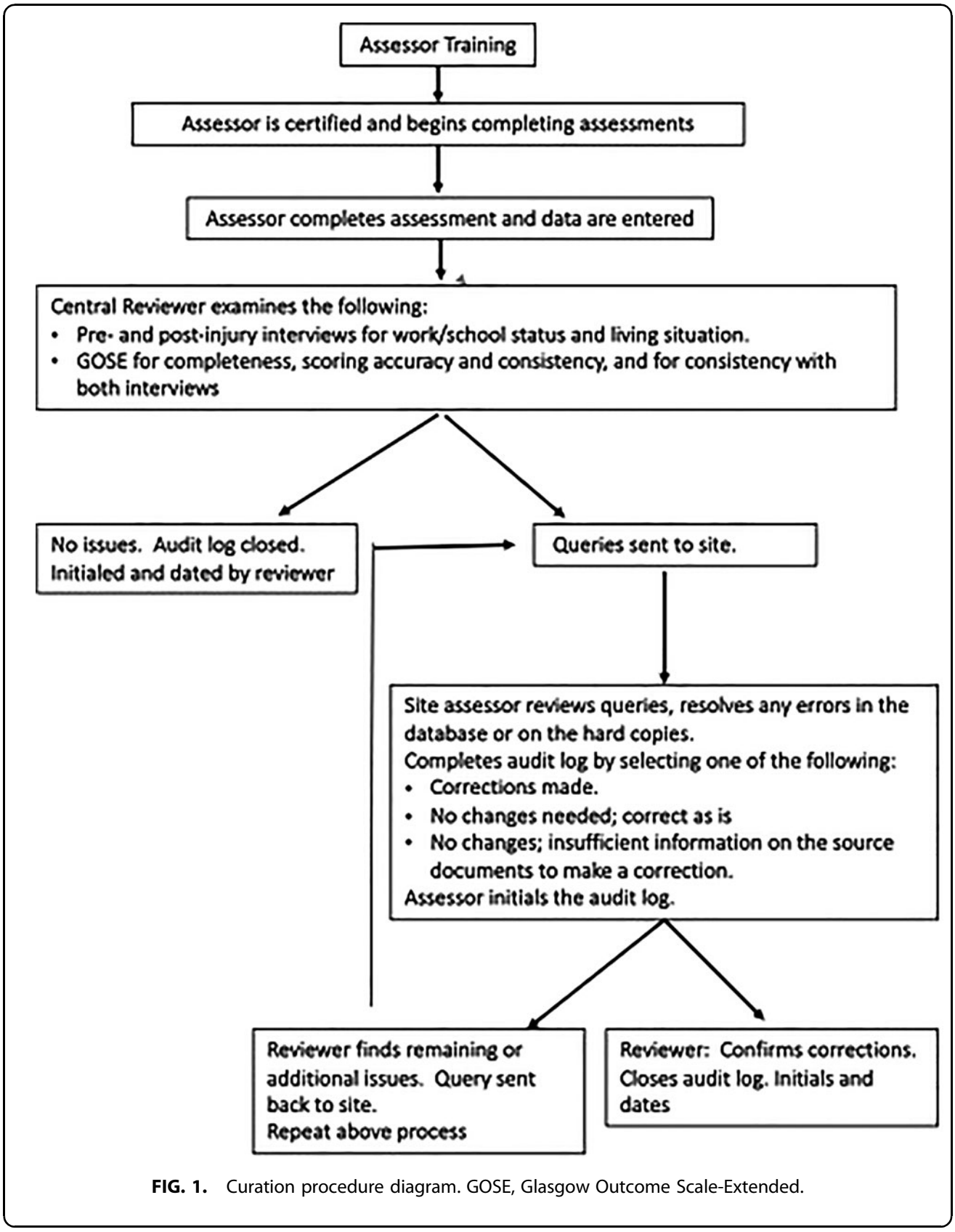

FIG. 1. Curation procedure diagram. GOSE, Glasgow Outcome Scale-Extended.

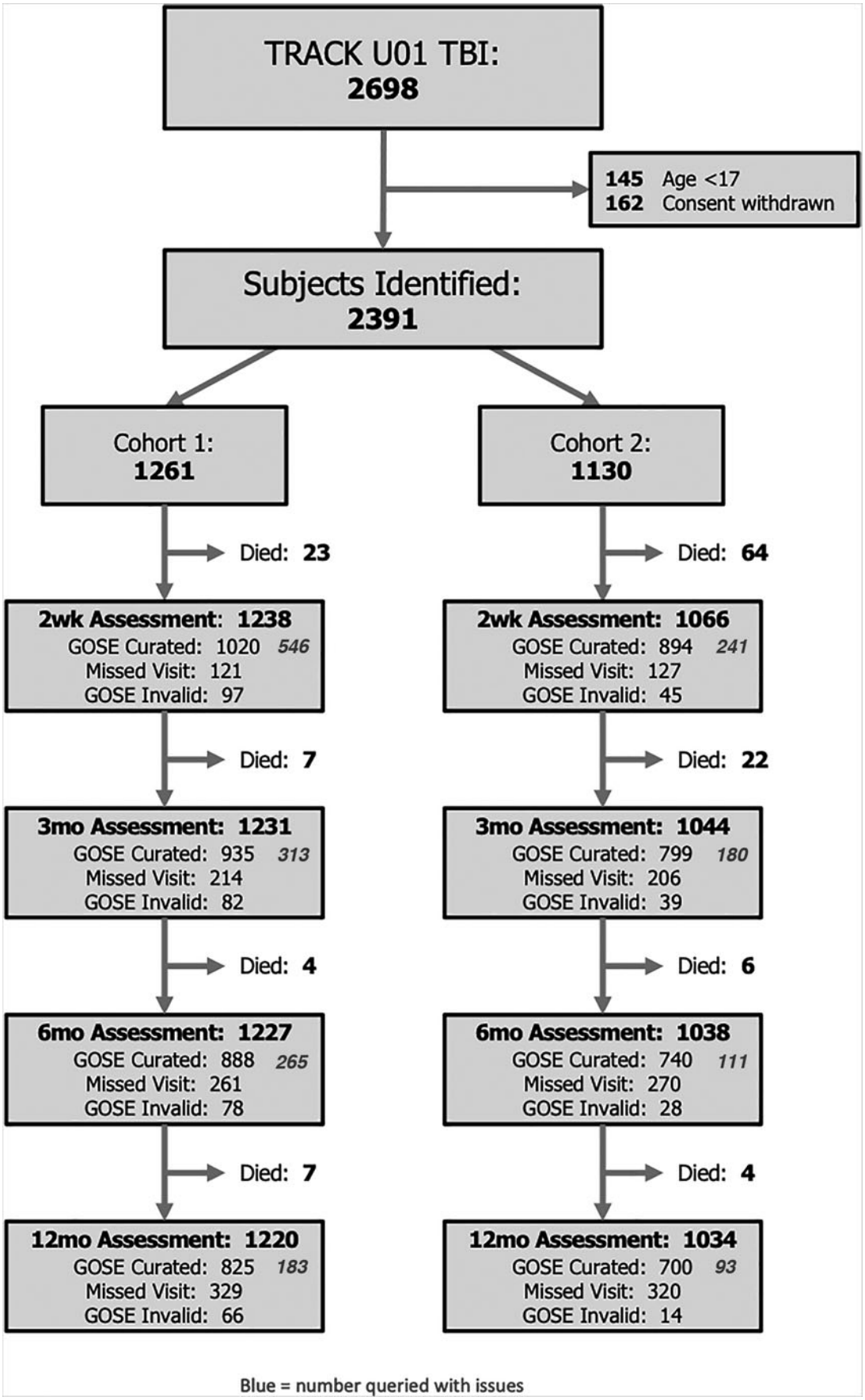

FIG. 2. Participant flow diagram: participants enrolled by cohorts 1 and 2. GOSE, Glasgow Outcome Scale-Extended.

Table 1. Participant Characteristics for TRACK-TBI GOSE Curation Cohort 1 and Cohort 2

	Cohort 1	*Cohort 2*
Participants with ≥1 GOSE reviewed	1261	1130
Age		
Mean (SD)	40.0 (16.7)	43.4 (18.2)
Sex		
Male	844 (67%)	808 (72%)
Education years		
Mean (SD)	13.4 (2.8)	13.2 (2.9)
Unknown	66	76
Injury cause		
Road traffic	769 (61%)	590 (53%)
Fall	298 (24%)	338 (30%)
Other accident	63 (5%)	61 (5%)
Violence	89 (7%)	77 (7%)
Other	41 (3%)	52 (5%)
Unknown	1	12
Glasgow Coma Scale on admission		
Mean (SD)	13.6 (3.3)	12.2 (4.3)
Median (IQR)	15 (14, 15)	15 (10, 15)
3–8, *N* (%)	120 (10)	229 (21)
9–12, *N* (%)	41 (3)	72 (7)
13–15, *N* (%)	1083 (87)	784 (72)
Unknown	17	45
Highest level of care		
Emergency department	376 (30%)	119 (11%)
Hospital, no intensive care unit	452 (36%)	362 (32%)
Intensive care unit	433 (34%)	649 (57%)

GOSE, Glasgow Outcome Scale-Extended; SD, standard deviation; IQR, interquartile range.

the weighted kappas were >0.90. These all fall within the range that is considered near-perfect agreement.[26]

Common reasons for Glasgow Outcome Scale-Extended corrections

Pre-injury status questions. Questions concerning how to rate change, if any, from pre-injury functional status or symptom burden proved especially difficult for assessors to master because of the GOSE's wording. For example, if the participant endorsed pre-injury relationship difficulties on Question 7 or having experienced other symptoms that affect daily life before the injury on Question 8, both common situations, the assessor was to ask whether those problems were worse now as compared with before the injury. If they were endorsed as ''worse now'' (indicating change since the injury), the SOP directed assessors to record a ''no'' response to having pre-injury difficulties. The GOSE's pre-injury status questions were changed in response to queries 466 times (some GOSEs had more than one instance) in cohort 1 and 62 times in cohort 2 (Table 3, panels A and B). The most common sections requiring changes to pre-injury status were: 1) ability to function at work and school ($n = 239$ corrections in cohort 1; $n = 35$ corrections in cohort 2); 2) level of relationship difficulties ($n = 89$ corrections in cohort 1; $n = 13$ corrections in cohort 2); and 3) return to normal life (symptom burden; $n = 75$ corrections in cohort 1; $n = 12$ corrections in cohort 2).

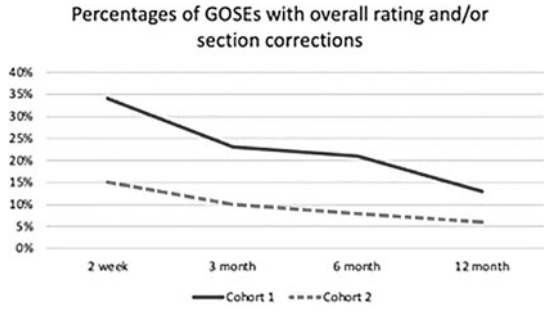

FIG. 3. Percentages of GOSE assessments with overall rating and/or section corrections by study cohort and follow-up time point. GOSE, Glasgow Outcome Scale-Extended.

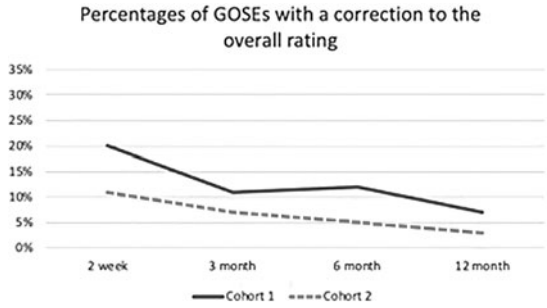

FIG. 4. Percentages of GOSE assessments with a correction to the overall rating by study cohort and follow-up time point. GOSE, Glasgow Outcome Scale-Extended.

Corrections to dependency questions. The dependency section ratings (assistance in the home, shopping, or travel) were corrected 163 times in cohort 1 and 30 times in cohort 2 (Table 4, panels A and B). Corrections occurred both for participants assessed when already at home not meeting criteria for dependency and those assessed while still hospitalized and not meeting criteria for independence.

Corrections to the level of disability in the Work/School section. In 72 cases in cohort 1 and 35 cases in cohort 2, the degree to which a participant was limited in their work capacity was corrected (Table 4, panels A and B). Almost all changed from a coding of reduced work capacity to being coded as unable to work. If physician clearance was required to return to work and this had not been granted, the participant was considered unable to work even if they believed they were able to work at least part time. This was the situation in nearly all cases where a correction was necessary.

Scoring errors. Scoring errors were defined as an overall rating that did not reflect the most severe limitation

Table 2. Degree of Change in GOSEs Post-Curation; Cohorts 1 and 2

Original score	Cohort 1:post-curation score															
	GOSE score: TBI rating								GOSE score:all rating							
2wk	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	22								22							
2		27			1					20						
3		1	115	3	11	2				1	99	1	10			
4			2	40	13	2		3			3	34	22	2		
5			3	1	153	6	3	2				2	160	1		
6					31	166	7	4					26	127	5	
7			1		8	9	193	3					4	8	80	5
8				1	7	5	1	205			1	1	3	6	4	76
	Unweighted kappa = 0.851;p<0.0001. Weighted kappa = 0.905 p<0.0001.								Unweighted kappa = 0.825;p<0.0001. Weighted kappa = 0.902;p<0.0001.							

Original score	Cohort 2:post-curation score															
	GOSE score: TBI rating								GOSE score: all rating							
2wk	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	63	1							63	1						
2		69								69	1					
3			153	7	3		2				198	3	1			
4			4	34	4	4		1			6	44	4	1		
5				3	187	7	4	2			2	1	267	2	2	1
6			1	1	10	141	1				1	1	15	121	1	
7			1	2	3	9	127	2				4	3	6	97	1
8				2	4	2	4	107				1	2	3	3	41
	Unweighted kappa = 0.898;p<0.0001. Weighted kappa = 0.939;p<0.0001.								Unweighted kappa = 0.917;p<0.0001. Weighted kappa = 0.951;p<0.0001.							

3mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	27								27							
2		4								3						
3			47	1		3	1	1			37			3		
4				19	7	1	2	1			3	21	5	1	1	1
5					106	2	5					1	122	2	2	
6					10	155	8						11	149	3	1
7				1	3	3	216	6					2	2	153	2
8					3	1		336				1	2	3	3	210
	Unweighted kappa = 0.921;p<0.0001. Weighted kappa = 0.943;p<0.0001.								Unweighted kappa = 0.920;p<0.0001. Weighted kappa = 0.947;p<0.0001.							

3mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	84								84							
2		10								10						
3			63	1	3						67	1	1			
4			3	25			2	1			3	31	4		1	1
5				1	116	2	3					1	162	1	1	
6					11	146	4	1					9	165	2	1
7					3	3	191	2					2	7	182	2
8					1		5	205					1	1	2	144
	Unweighted kappa = 0.937;p<0.0001. Weighted kappa = 0.968;p<0.0001.								Unweighted kappa = 0.944;p<0.0001. Weighted kappa = 0.973;p<0.0001.							

6mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	31								31							
2		2								1						
3			38	1	1	1					33		1			
4				10	2	2	1					16	1	3	1	
5				1	67	2	5	1				1	81	1	4	
6					9	135	11	2					13	162	7	1
7					1	6	238	12					2	8	190	7
8					1	1	4	335			1			3	10	294
	Unweighted kappa = 0.906;p<0.0001. Weighted kappa = 0.945;p<0.0001.								Unweighted kappa = 0.904;p<0.0001. Weighted kappa = 0.944;p<0.0001.							

6mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	90								90							
2		4								4						
3			48			1		1			53					
4			1	11	1	1					1	13	1			
5			1		73	4	2	1					97	4		
6					5	147	4						6	167	4	1
7						2	199	2						4	193	1
8							1	229					1	1	1	187
	Unweighted kappa = 0.959;p<0.0001. Weighted kappa = 0.980;p<0.0001.								Unweighted kappa = 0.963;p<0.0001. Weighted kappa = 0.985;p<0.0001.							

12mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	38								38							
2																
3			25		1						31					
4				15	4		2					13	4		2	
5					56	6	2	4				1	76	3	2	1
6					5	115	5	2				1	4	131	3	
7						7	191	5				1		5	180	5
8				1	2			378					2	1	3	357
	Unweighted kappa = 0.926;p<0.0001. Weighted kappa = 0.952;p<0.0001.								Unweighted kappa = 0.940;p<0..0001. Weighted kappa = 0.965;p<0..0001.							

12mo	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	93								93							
2		2								2						
3			33								35					
4			1	16				1			1	16				1
5					62	3	1	1					76	2	1	
6						125	2						2	128	1	
7					1	1	175	5						2	185	4
8					1	1	2	249					2	1	1	222
	Unweighted kappa = 0.967;p<0.0001. Weighted kappa = 0.983;p<0.0001.								Unweighted kappa = 0.971;p<0.0001. Weighted kappa = 0.985;p<0.0001.							

GOSE, Glasgow Outcome Scale-Extended.

Table 3. Number of Corrections Made to Questions about Pre-Injury Limitations

A. Cohort 1	*Total (n = 3668)*	*2 week (n = 1020)*	*3 month (n = 935)*	*6 month (n = 888)*	*12 month (n = 825)*
Pre-injury assistance in home	6	3	2	1	0
Pre-injury unable to shop	15	5	8	2	0
Pre-injury unable to travel	9	4	2	3	0
Pre-injury work/school	239	77	74	54	34
Pre-injury social and leisure activity	33	9	11	11	2
Pre-injury relationship discord	89	27	16	35	11
Pre-injury symptoms	75	28	15	19	13
B. Cohort 2	*Total (n = 3133)*	*2 week (n = 894)*	*3 month (n = 799)*	*6 month (n = 740)*	*12 month (n = 700)*
Pre-injury assistance in home	0	0	0	0	0
Pre-injury unable to shop	0	0	0	0	0
Pre-injury unable to travel	0	0	0	0	0
Pre-injury work/school	35	19	5	8	3
Pre-injury social and leisure activity	2	0	2	0	0
Pre-injury relationship discord	13	2	2	5	4
Pre-injury symptoms	12	1	4	6	1

Panel A summarizes findings for cohort 1; panel B summarizes cohort 2.

represented by a change from pre-injury status; this occurred in 196 cases (5%) in cohort 1 and in 165 cases (5%) in cohort 2.

Insufficient information available to justify making a rating change. In 45 cases in cohort 1 and five cases in cohort 2, site assessors determined that there was insufficient information recorded on the source documents to justify changing a rating. Some of these inconsistencies could not be reconciled because the review was done too long after the assessment was completed or the original assessor had left the project and had not documented the situation with enough clarity.

Evaluation of potential confounding factors

Although we suspected that decreased queries and corrections to GOSE entries over time was attributable to ongoing curation and training, at least three confounding factors could have also contributed to different trends in queries. First, we considered that the shift from the simpler GOSE-TBI to the more complex GOSE 2-ways interview might lead to more queries when the GOSE 2-ways interview was introduced early in the study. However, the percentage of corrections was similar for the two interviews (29% at 2 weeks for the GOSE-TBI and 32% for the GOSE 2-ways interview), implying that this change had minimal impact on the percentage of corrections required (Supplementary Table S2). Second, we considered that the shift in enrollment priorities to more-severe TBI cases in cohort 2 could have explained the decrease in corrections from cohort 1 to cohort 2 if the more severely injured cases were easier to rate.

Contrary to this hypothesis, we observed substantial decreases in error corrections within each TBI severity group from cohort 1 to cohort 2 and similar rates of corrections across the different severity levels (Supplementary Table S3). Third, we examined the possibility that

Table 4. Number of Corrections to the Degree of Dependency or Degree of Work Limitations

A. Cohort 1 Corrections to dependency questions	*Total (n = 3668)*	*2 week (n = 1020)*	*3 month (n = 935)*	*6 month (n = 888)*	*12 month (n = 825)*
Assistance in the home	58	37	9	7	5
Ability to shop	60	29	14	7	10
Ability to travel	45	23	8	7	7
Work/school	*Total*	*2 week*	*3 month*	*6 month*	*12 month*
Work ability level corrections (reduced/unable; primarily physician clearance)	72	38	20	9	5
B. Cohort 2 Corrections to dependency questions	*Total (n = 3133)*	*2 week (n = 894)*	*3 month (n = 799)*	*6 month (n = 740)*	*12 month (n = 700)*
Assistance in the home	13	4	7	0	2
Ability to shop	10	3	5	0	2
Ability to travel	7	5	1	1	0
Work/school	*Total*	*2 week*	*3 month*	*6 month*	*12 month*
Work ability level corrections (reduced/unable; primarily physician clearance)	35	16	11	5	3

error percentages were skewed by a few sites or assessors. Inconsistences and errors were noted across all sites (data not shown). For example, in cohort 1 at the 2-week follow-up, query percentages ranged from 34% to 75% by site. Corrections to the GOSE ranged from 20% to 45% by site. Although there was variation across sites, scoring difficulties of some level were observed at all sites.

Discussion

In this comparative analysis of 6801 GOSE interviews collected at 18 TRACK-TBI study sites over two phases of the study, site assessors' ratings and rating practices improved substantially after central review. Routine audit, feedback, and retraining resulted in substantially fewer errors and rarer need for recoding of the GOSE, resulting in more reliable data for this widely used measure of functional outcome after TBI.

Wilson and colleagues' (2007) study described concerning variances in GOSE scoring, then showed a decrease in variability after a period of review and feedback completed by central review. The rate of inconsistency we observed in overall ratings in the cohort 1 (e.g., 20% at 2 weeks post-injury) period fell consistently in the range of previous reviews from 17%[13] to 40%.[11] With retraining, review, and feedback, the frequency of corrections in 2-week post-injury overall ratings dropped to 11% in cohort 2.

The curation process evolved over a period of time as inconsistencies between assessors and sites became apparent. The final curation process, as described here, resulted from recognition that errors and inconsistencies were not being captured and addressed by the training and data review systems we originally set in place.

Lessons learned from central Glasgow Outcome Scale-Extended reviews

Benefit of central review. The queries identified discrepancies, described the type of information one might need to consider, or referenced pre-injury information which may have been missed. Over time, assessors began to think in those terms. Issues observed in determining dependence inside and outside of the home were almost non-existent by the time the central reviewers analyzed the cohort 2 data. Those that did remain were likely borderline cases that will always be especially difficult to interpret and score. Corrections to the work section also decreased dramatically as assessors developed deeper appreciation of the need to review the pre-injury employment/school status of each participant (e.g., those who were identified within the work group and students before injury), as well as better understanding of changes in participants' ability to execute work or academic requirements.

Ongoing monitoring to ensure that assessors are following suggested guidance is also important. Some studies, such as BOOST-3,[27] have a small number of highly trained assessors to administer the GOSE for all participants in a multi-site study to maximize consistency. Box 3 summarizes the guidance provided to TRACK-TBI outcome assessors.

Study-specific decisions. An advantage of the GOSE is its flexibility to adjust the measure to best capture questions unique to each project. Instances of this flexibility include the decision as to whether or not to include effects of peripheral injuries in the ratings or include passive activities (as opposed to only activities outside the home) in rating the social and leisure section. Some studies may include caregiving in the work category or define work in another way (although this was not done in the present study). This same flexibility can add to variation in scoring the measure and, consequently, differences in the interpretation of scores across studies. It is important that such questions and decisions be addressed at the beginning of the study or as soon as they are recognized in order to obtain accurate, consistent scoring. Whatever the decisions, it is important that information is collected and coded consistently across time and research sites. It is also important that investigators report the method used to score the GOSE to enable the comparison of results across studies.

Interview the best source. The GOSE can be administered to the participant, a family member, friend, or caregiver. Sometimes, information in the medical record is sufficient to code or clarify the overall score. For instance, for those most severely injured and still hospitalized, the medical record can sometimes provide sufficient information to assign an overall rating in the vegetative or severe disability categories (ratings of 2, 3, and, at times, 4). Instructions are to use the best source of information, which, in most cases, is the participant. However, this is not always the case. For example, a note written by an assessor indicated that a hospitalized participant answered that he was able to shop and travel independently and the assessor accordingly coded him as being independent. However, the assessor observed that he wore a Wander Guard to alert staff if he was attempting to leave the unit. The participant may have felt capable in this regard, but a family member or care team member might have given a more accurate account of his current level of ability; the Wander Guard being evidence that he was not cleared to be independent by his medical care team. Some patients may overestimate their abilities or simply lack awareness of their limitations. Assessors should be alert to this and investigate further, if necessary.

Box 3. Examples of Training Guidance for Interpretation of GOSE Sections

GOSE Section	*Considerations*	*TRACK-TBI Guidance*
Assistance in the home	Determining when assistance becomes dependence	Focus on dependence and safety; minimize the emphasis on assistance. Emphasize that this is a very basic level of care. Assistance with 1 or 2 activities such as washing hair or tying shoelaces, showering, should not be scored as dependent.
	Participants who are hospitalized	Being hospitalized or in a care facility by definition required a score of some level of dependence. TRACK encouraged assessors to ask questions about what a participant was able to do on their own with no one else in the room. Were they cleared to get up and use the restroom without supervision? Manage basic ADLs? If they were able to do these things they generally achieved a rating of 4 on GOSE-All. If the primary reason for hospitalization was because of non-brain injuries, respondents were asked what they thought they would be able to do if they did not have the non-brain injuries.
	Participants with orthopedic casts	For the GOSE-All rating assessors were reminded to use the same criteria – focus on dependence and safety – not just assistance.
Shopping and Travel	Participants who are hospitalized	Assessors were encouraged to ask if a participant was cleared to leave the unit by themselves to go, for example, to the cafeteria or gift shop. Are they cleared to leave the facility by themselves? If they did not have medical clearance they were considered dependent in these areas, GOSE rating of 4.
	Defining what constitutes shopping or travel	Shopping independence: The ability to purchase one item or order from a menu. Online shopping was not considered since the section is meant to assess independence outside the home. Travel independence: Able to go places away from home without supervision. Calling a friend or a taxi for transportation was acceptable, but to be scored as independent required their ability to proceed on their own once they got out of the vehicle.
Work/School	Defining work; set parameters for applicability	Work was deemed applicable if the participant was not a homemaker, retired, or permanently disabled. Also, see school below.
	School	Reminders were given that this section applies to students. A second line of questioning pertaining to the ability to return to school at their prior level was provided.
	Participants who need physician clearance to return to work	If a participant needed physician clearance to return to their job and had not received this, they were considered unable to work (rating of 5) regardless of whether they thought they could do their job or not.
	Assess in a consistent way across time	Assessors were encouraged to note the pre-injury work status of each participant before an upcoming follow-up. Were they employed? In school? Reconcile any conflicts early on and proceed accordingly.
	Participants in the role of both worker and student who have returned to one activity but not the other.	As the GOSE looks at change from pre- to post-injury, the ability to return to one activity but not both constituted, in general, a reduction in work/school capacity.
Social and Leisure	Rating passive activities within the home	Although the structured interview asks about social and leisure activities outside the home, TRACK-TBI included passive activities within the home as well as those outside the home. By broadening the area, assessors were encouraged to find some activity for every participant. Any activity done to relax or for enjoyment was counted.
Relationships	Separating relationship difficulties from the emotional symptoms that cause them	Emphasis in the initial feedback and training from cohort 1 was focused on clarifying that issues with relationships were rated, as opposed to the emotional difficulties that can cause them; e.g., an outcomes assessor's notes written might record: 'Participant has a great deal of anxiety' which would generate a reminder query to focus on the impact of anxiety on relationships.
The GOSE in general	Participants who sustain a new injury	As much as possible the TBI rating was meant to capture just the effect of the index TBI. All new injury difficulties were included in the 'All' rating even, if it included a new TBI. Notes were written and the new injury was documented. As much as possible the TRACK guidance was to include difficulties from injuries not illnesses.
	Participants with pre-injury difficulties or disabilities	Queries were sent each time a participant was coded as having pre-injury difficulties to double check data entry errors.
	Unusual or complicated situations	Assessors were encouraged to document any unusual situations. They were encouraged to present difficult-to-code scenarios on monthly conference calls to develop consensus with the group. Built into the database was a 'confounding issues' text field where assessors could note unusual circumstances.
	Complete the full interview if not vegetative (rating of 2)	The entire interview was given. In the event an assessor learned new information that altered the way a prior section was coded they were advised to go back to that section and query further, changing a rating if necessary.
	Review corrections	It was not uncommon for an assessor to make a correction based upon a query only to then have it create another error, many times failing to reassign a new overall rating. It was important for the central review team to review a case after changes were made in response to queries.

Consistency with ambiguous situations. The GOSE is typically used to assess the independence of persons in the community. Because some participants in TRACK-TBI were still hospitalized for TBI or other system injuries at the time of the first follow-up at 2 weeks, some guidance was needed to help assessors determine the rating. The fact that they were hospitalized or in some other care facility deemed them dependent, warranting a rating of 3 or 4 (lower or upper severe disability), at least for GOSE-All. Assessors queried the kinds of activities the participant was completing on their own and those they required assistance with. Assessors were encouraged to ask the participant if they could go to the cafeteria or gift shop alone (independent in shopping), or if they needed someone with them. Another clarifying query was whether the participant was cleared to leave the hospital on their own (independent in travel).

Priorities for Glasgow Outcome Scale-Extended assessor training

The second goal of this work was to identify the sections within the GOSE that were associated with the most variation in scoring. Our results highlighted particular sections of the GOSE interview that are prone to administration and scoring errors. Below, we present priority areas for focused training efforts.

Challenges interpreting post-injury function relative to pre-injury function. Each section of the GOSE contains a question referencing difficulties in the particular area of function before the injury (e.g., ''Did the participant experience relationship difficulties prior to the injury?''). The interpretation of answers to the ''pre-injury'' questions resulted in the greatest degree of correction. In keeping with standard GOSE administration protocol, we trained site assessors to determine whether a subject's endorsement of these questions constituted a change, that is, endorsing the question suggests that there was not a change from pre- to post-injury in that particular area of function. Accordingly, the assessor was to discount that question in determining the overall rating.

Pre-injury difficulty with relationships ($n=89$ corrections in cohort 1; $n=13$ in cohort 2). Many participants endorsed having had relationship difficulties before the injury. According to the GOSE interview protocol, if a participant had relationship difficulties before the injury but those difficulties were worse after the injury, the question addressing pre-injury difficulty needed to be coded to indicate that a change had occurred, that is by indicating that the participant did not have relationship difficulties (to this degree) before the injury. This caused confusion given that the assessor had to record the opposite of the participant's response (see section 6.8, specifically Q7c of *A Manual for the GOSE Interview*).[14]

Pre-injury symptoms ($n=75$ corrections in cohort 1; $n=12$ in cohort 2). Patients experience symptoms, such as headaches, even before sustaining a TBI. Thus, when asked the question, ''Were similar symptoms present before the injury?,'' many respondents replied ''yes.'' However, if the symptoms were worse now, we directed that this question be coded ''no,'' to indicate a change from pre- to post-injury, again using the GOSE protocol (see section 6.9, specifically Q8b of *A Manual for the GOSE Interview*).[14]

Pre-injury work/school status ($n=239$ corrections in cohort 1; $n=35$ corrections in cohort 2). TRACK-TBI determined the work section of the GOSE that was to be completed for anyone who was not fully retired, permanently disabled, or a homemaker before the injury. This section also applies to students, which assessors sometimes forgot. It was often erroneously not administered when a participant had been unemployed before the injury. If the participant was unemployed before the injury, but medically cleared to work, the established procedure was for the work questions to be asked as hypotheticals. There were also times it was administered when it need not have been, as in the case of someone fully retired, disabled, or a homemaker before the injury.

Pre-injury dependency questions: assistance in the home, shopping, and travel ($n=30$ corrections in cohort 1; and none in cohort 2). The dependency pre-injury questions can be difficult to assess, given that it is often impossible to determine exactly how much assistance someone who was aged or disabled actually needed. Assessors relied on notes describing the pre-injury situation, the pre-injury interview, and responses to other sections of the GOSE to make final determinations regarding whether there was a change from the pre-injury level of dependence.

Understanding what each Glasgow Outcome Scale-Extended section assesses. Assessors also required additional instruction to understand what several sections of the GOSE interview are intended to assess, as illustrated by the following examples.

Assistance in the home. The categories assessing dependence in the home and the ability to shop and travel were the most problematic. Even with a Manual and operational definitions, assessors interpreted these limitations in many different ways, specifically, the point at which ''assistance'' becomes dependence. The criterion used to rate this category is whether or not the person can manage safely on their own for a 24-h period if necessary (see Box 2, severe disability [SD] of the GOSE administration manual).[14] Many times, a participant responded ''yes'' to getting assistance only to indicate when queried further that it was for higher-level tasks, such as showering or cooking, or for one non-vital

activity, but was otherwise independent and safe. In these cases, it was important that such a person not be considered severely disabled. Orthopedic casts influenced the ratings in this area greatly. Casts are cumbersome and assistance is often forthcoming for activities such as showering and dressing. Except in extreme circumstances (e.g., non-weight-bearing) casts alone would generally not place someone in a category of dependency. In a clinical trial, where a dichotomized GOSE is used, describing a poor outcome as a score from 1 to 4 and a good outcome as a score from 5 to 8, the effect could be substantial.

Independence in shopping. All that is required to be considered independent in this area is for a participant to be able to make a small purchase. This caused confusion among assessors, such as one who indicated that a participant was unable to shop because he could not carry all of the shopping bags. The GOSE is designed to be a semistructured measure allowing for latitude in how the questions are asked, but always keeping in mind the goal of a particular section.

Glasgow Outcome Scale-Extended corrections influenced by medical clearance. Medical clearance takes precedence when applicable. Corrections to the dependency questions occurred in cases of participants still hospitalized. The participant may have felt capable of being on their own or being able to travel independently, but being hospitalized deemed them dependent (an overall rating of 3 or 4), at least on the GOSE-All.

Medical clearance also impacted the work/school section. Almost all corrections in the ability to return to work or school ($n = 72$ corrections in cohort 1; $n = 35$ in cohort 2) resulted from being coded as able to work in a reduced capacity to being coded as unable to work. Many times, a participant reported that they felt capable of working part time. However, if they needed physician clearance to return and this had not been granted, they were considered unable to work in this study.

Future studies will benefit from the newly created GOSE manual, *A Manual for the Glasgow Outcome Scale-Extended (GOSE) Interview* (Wilson and colleagues). The Manual will facilitate the training of assessors and help to maintain standardization.

Strengths and limitations of this study

The large number of GOSE interviews reviewed, broad range of functional impairment within the sample, and large number of sites and assessors are factors that facilitate the generalizability of these findings. The curation process was led by a small group (K.B., J.M., and G.S.), all very experienced in the administration of the GOSE. Some curation had already taken place before the decision was made to review all GOSEs and track queries and changes. Therefore, the review numbers indicated above are likely underestimates of the errors on the GOSEs, with the underestimation being greater in cohort 1. Finally, as is customary for large multi-center studies, assessors were not directly observed conducting participant interviews by the central training staff, and curation efforts were limited to review of entered data for logical inconsistencies. It is possible that direct observation or review of audio recordings of interviews would reveal additional insights into rater errors and ways to improve training and quality assurance activities for studies using the GOSE interview.

Conclusion

Many lessons have been learned over the course of the TRACK-TBI study. The GOSE is replete with nuance, making it important for investigators to thoughtfully implement rating decisions, and provide training with added focus to the difficult aspects of the interview. Ongoing monitoring and timely curation to promote accuracy and consistency should not be overlooked.

The advantages of the GOSE are its brevity, universality, and utility in documenting disability as the result of TBI. Though the GOSE is not difficult to administer, it certainly is more difficult to learn the inherent nuances than was previously recognized. It is imperative that assessors fully understand the purpose of each section within the interview.

Although TRACK-TBI trained and certified assessors, 24% of administrations required correction before the institution of contemporaneous curation. Had the TRACK-TBI outcomes team been aware of the specific difficulties experienced by assessors at the beginning of the study, a great deal of time would have been saved in the curation process. By cohort 2, the time spent reviewing data was greatly decreased. Whereas many of the lessons learned applied to the GOSE, such as the importance of assessor training, data quality checks, and central curation, they could certainly be applied to other outcome measures and other studies. The GOSE is not unique in regard for the need of training and monitoring. The results of this curation effort highlight the importance of having a central oversight review team both to identify errors and also help ensure consistent administration of outcome measures across multiple assessors and sites. Errors increase the variability in outcomes and decrease the ability of a study to detect real effects. Avoiding differences across sites in the interpretation of GOSE sections is especially important in observational comparative effectiveness studies. Establishing clear scoring rules and providing ongoing guidance and feedback to data collectors are essential to avoid misclassification of the GOSE. Many TRACK-TBI assessors worked on multiple projects involving the GOSE. Because it can be administered in different ways, it is important that individual studies review data to ensure consistency within their projects.

The GOSE continues to be the most commonly used and, in the case of clinical trials, often the primary clinical outcome measure of TBI studies. The results presented herein indicate that, particularly in the setting of multi-site studies, the GOSE requires extensive training, ongoing monitoring, timely curation, and central oversight review to improve accuracy of administration and scoring, and these are necessary to optimize its sensitivity as a primary end-point.

The TRACK-TBI Investigators

Opeolu Adeoye, MD, University of Cincinnati; Neeraj Badjatia, MD, University of Maryland; M. Ross Bullock, MD, PhD, University of Miami; Randall Chesnut, MD, University of Washington; John D. Corrigan, PhD, ABPP, Ohio State University; Karen Crawford, University of Southern California; Ramon Diaz-Arrastia, MD, PhD, University of Pennsylvania; Ann-Christine Duhaime, MD, MassGeneral Hospital for Children; Richard Ellenbogen, MD, University of Washington; V Ramana Feeser, MD, Virginia Commonwealth University; Adam R. Ferguson, PhD, University of California, San Francisco; Brandon Foreman, MD, University of Cincinnati; Raquel Gardner, University of California, San Francisco; Etienne Gaudette, PhD, University of Southern California; Dana Goldman, PhD, University of Southern California; Luis Gonzalez, TIRR Memorial Hermann; Shankar Gopinath, MD, Baylor College of Medicine; Rao Gullapalli, PhD, University of Maryland; J Claude Hemphill, MD, University of California, San Francisco; Gillian Hotz, PhD, University of Miami; Sonia Jain, PhD, University of California, San Diego; C. Dirk Keene, MD, PhD, University of Washington; Frederick K. Korley, MD, PhD, University of Michigan; Joel Kramer, PsyD, University of California, San Francisco; Natalie Kreitzer, MD, University of Cincinnati; Harvey Levin, MD, Baylor College of Medicine; Chris Lindsell, PhD, Vanderbilt University; Christopher Madden, MD, UT Southwestern; Alastair Martin, PhD, University of California, San Francisco; Thomas McAllister, MD, Indiana University; Randall Merchant, PhD, Virginia Commonwealth University; Pratik Mukherjee, MD, PhD, University of California, San Francisco; Laura B. Ngwenya, MD, PhD, University of Cincinnati; Florence Noel, PhD, Baylor College of Medicine; Amber Nolan, MD, PhD, University of California, San Francisco; David Okonkwo, MD, PhD, University of Pittsburgh; Eva Palacios, PhD, University of California, San Francisco; Daniel Perl, MD, Uniformed Services University; Ava Puccio, PhD, University of Pittsburgh; Miri Rabinowitz, PhD, University of Pittsburgh; Claudia Robertson, MD, Baylor College of Medicine; Jonathan Rosand, MD, MSc, Massachusetts General Hospital; Angelle Sander, PhD, Baylor College of Medicine; David Schnyer, PhD, UT Austin; Seth Seabury, PhD, University of Southern California; Mark Sherer, PhD, TIRR Memorial Hermann; Arthur Toga, PhD, University of Southern California; Alex Valadka, MD, Virginia Commonwealth University; Mary Vassar, RN MS, University of California, San Francisco; Paul Vespa, MD, University of California, Los Angeles; Kevin Wang, PhD, University of Florida; John K. Yue, MD, University of California, San Francisco; Esther Yuh, MD, PhD, University of California, San Francisco; Ross Zafonte, Harvard Medical School.

Funding Information

This work was supported by NIH-NINDS – TRACK-TBI (Grant #U01NS086090) and United States Department of Defense – TBI Endpoints Development Initiative (Grant #W81XWH-14-2-0176). Dr. Manley discloses grants from the United States Department of Defense – TBI Endpoints Development Initiative (Grant #W81XWH-14-2-0176), TRACK-TBI Precision Medicine (Grant # W81XWH-18-2-0042), and TRACK-TBI NETWORK (Grant # W81XWH-15-9-0001); NIH-NINDS – TRACK-TBI (Grant #U01NS086090); and the National Football League (NFL) Scientific Advisory Board – TRACK-TBI LONGITUDINAL. United States Department of Energy supports Dr. Manley for a precision medicine collaboration. One Mind has provided funding for TRACK-TBI patients stipends and support to clinical sites. He has received an unrestricted gift from the NFL to the UCSF Foundation to support research efforts of the TRACK-TBI NETWORK. Dr. Manley has also received funding from NeuroTruama Sciences LLC to support TRACK-TBI data curation efforts. Additionally, Abbott Laboratories has provided funding for add-in TRACK-TBI clinical studies.

Amy Markowitz receives funding from the Department of Defense TBI Endpoints Development Initiative (Grant #W81XWH-14-2-0176) and TRACK-TBI NETWORK (Grant # W81XWH-15-9-0001). Ms Markowitz also receives salary support from the United States Department of Energy precision medicine collaboration and the philanthropic organization, One Mind.

Lindsay Wilson is supported by CENTER-TBI (EU FP7 programme 602150). Lindsay Nelson is supported by NINDS grant R01 NS110856.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Box S1
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3

References

1. Jennett, B., and Bond, M. (1975). Assessment of outcome after severe brain damage. Lancet 1, 480–484.
2. Wilson, J.T., Pettigrew, L.E., and Teasdale, G.M. (1998). Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. J. Neurotrauma 15, 573–585.

3. McMillan, T., Wilson, L., Ponsford, J., Levin, H., Teasdale, G., and Bond, M. (2016). The Glasgow Outcome Scale - 40 years of application and refinement. Nat. Rev. Neurol. 12, 477–485.
4. Wilde, E.A., Whiteneck, G.G., Bogner, J., Bushnik, T., Cifu, D.X., Dikmen, S., French, L., Giacino, J.T., Hart, T., Malec, J.F., Millis, S.R., Novack, T.A., Sherer, M., Tulsky, D.S., Vanderploeg, R.D., and von Steinbuechel, N. (2010). Recommendations for the use of common outcome measures in traumatic brain injury research. Arch. Phys. Med. Rehabil. 91, 1650–1660.e17.
5. National Institutes of Neurologic Disease and Stroke Common Data Elements. (2010). Traumatic brain injury: data standards. commondataelements.ninds.nih.gov/traumatic%20brain%20injury (Last accessed April 12, 2021).
6. Anderson, S., Housley, A., Jones, P., Slattery, J., and Miller, J. (1993). Glasgow Outcome Scale: an inter-rater reliability study. Brain Inj. 7, 309–317.
7. Brooks, D.N., Hosie, J., Bond, M.R., Jennett, B., and Aughton, M. (1986). Cognitive sequelae of severe head injury in relation to the Glasgow Outcome Scale. J. Neurol. Neurosurg. Psychiatry 49, 549–553.
8. Choi, S.C., Clifton, G.L., Marmarou, A., and Miller, E.R. (2002). Misclassification and treatment effect on primary outcome measures in clinical trials of severe neurotrauma. J. Neurotrauma 19, 17–22.
9. Maas, A.I., Braakman, R., Schouten, H.J., Minderhoud, J.M., and van Zomeren, A.H. (1983). Agreement between physicians on assessment of outcome following severe head injury. J. Neurosurg. 58, 321–325.
10. Teasdale, G.M., Pettigrew, L.E., Wilson, J.T., Murray, G., and Jennett, B. (1998). Analyzing outcome of treatment of severe head injury: a review and update on advancing the use of the Glasgow Outcome Scale. J. Neurotrauma 15, 587–597.
11. Wilson, J.T., Slieker, F.J., Legrand, V., Murray, G., Stocchetti, N., and Maas, A.I. (2007). Observer variation in the assessment of outcome in traumatic brain injury: experience from a multicenter, international randomized clinical trial. Neurosurgery 61, 123–128; discussion, 128–129.
12. van Baalen, B., Odding, E., van Woensel, M.P., and Roebroeck, M.E. (2006). Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. Clin. Rehabil. 20, 686–700.
13. Marmarou, A. (2001). *Head Trauma: Basic, Preclinical, Clinical Direction*. Wiley: New York.
14. Wilson, L., Boase, K., Nelson, L.D., Temkin, N.R., Giacino, J.T., Markowitz, A.J., Maas, A., Menon, D., Teasdale, G., and Manley, G.T. (2021). A Manual for the Glasgow Outcome Scale-Extended (GOSE) Interview. J. Neurotrauma. 38, 2435–2446.
15. Lu, J., Murray, G.D., Steyerberg, E.W., Butcher, I., McHugh, G.S., Lingsma, H., Mushkudiani, N., Choi, S., Maas, A.I., and Marmarou, A. (2008). Effects of Glasgow Outcome Scale misclassification on traumatic brain injury clinical trials. J. Neurotrauma 25, 641–651.
16. Maas, A.I., Steyerberg, E.W., Marmarou, A., McHugh, G.S., Lingsma, H.F., Butcher, I., Lu, J., Weir, J., Roozenbeek, B., and Murray, G.D. (2010). IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. Neurotherapeutics 7, 127–134.
17. Horton, L., Rhodes, J., and Wilson, L. (2018). Randomized controlled trials in adult traumatic brain injury: a systematic review on the use and reporting of clinical outcome assessments. J. Neurotrauma 35, 2005–2014.
18. Ercole, A., Brinck, V., George, P., Hicks, R., Huijben, J., Jarrett, M., Vassar, M., and Wilson, L.; the DAQCORD collaborators. (2020). Guidelines for Data Acquisition, Quality and Curation for Observational Research Designs (DAQCORD). J. Clin. Transl. Sci 4, 354–359.
19. Lu, J., Marmarou, A., Lapane, K., Turf, E., and Wilson, L.; on behalf of the IMPACT Group and American Brain Injury Consortium Study Participation Centers. (2010). A method for reducing misclassification in the extended Glasgow Outcome Score. J. Neurotrauma 27, 843–852.
20. University of California, San Francisco. (2014). TRACK-TBI; Transforming Research and Clinical Knowledge in TBI. tracktbi.ucsf.edu (Last accessed April 12, 2021).
21. Giacino, J.T., Kalmar, K., and Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. Arch. Phys. Med. Rehabil. 85, 2020–2029.
22. Temkin, N.R., Zahniser, E., Morrissey, M.R., Barber, J., Machamer, J., Manley, G., and Dikmen, S. (2018). Glasgow Outcome Scale Extended-disability from only TBI vs all injuries in those with GCS 13–15: a TRACK-TBI study. J. Neurotrauma 2018;35: A-67–A-68.
23. Maas, A.I., Menon, D.K., Steyerberg, E.W., Citerio, G., Lecky, F., Manley, G.T., Hill, S., Legrand, V., and Sorgner, A.; CENTER-TBI Participants and Investigators. (2015). Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. Neurosurgery 76, 67–80.
24. TRACK-TBI; International Traumatic Brain Injury Initiative. (2018). Standard Operating Procedures for Outcome Assessment Version 10. tracktbi.ucsf.edu/sites/tracktbi.ucsf.edu/files/Outcome%20Assessment%20SOP_04-12-18_V10_clean.pdf (Last accessed April 12, 2021).
25. Nelson, L.D., Temkin, N.R., Dikmen, S., Barber, J., Giacino, J.T., Yuh, E., Levin, H.S., McCrea, M.A., Stein, M.B., Mukherjee, P., Okonkwo, D.O., Diaz-Arrastia, R., Manley, G.T., and the TRACK-TBI Investigators; Adeoye, O., Badjatia, N., Boase, K., Bodien, Y., Bullock, M.R., Chesnut, R., Corrigan, J.D., Crawford, K., MIS; Duhaime, A.C., Ellenbogen, R., Feeser, V.R., Ferguson, A., Foreman, B., Gardner, R., Gaudette, E., Gonzalez, L., Gopinath, S., Gullapalli, R., Hemphill, J.C., Hotz, G., Jain, S., Korley, F., Kramer, J., Kreitzer, N., Lindsell, C., Machamer, J., Madden, C., Martin, A., McAllister, T., Merchant, R., Noel, F., Palacios, E., Perl, D., Puccio, A., Rabinowitz, M., Robertson, C.S., Rosand, J., Sander, A., Satris, G., Schnyer, D., Seabury, S., Sherer, M., Taylor, S., Toga, A., Valadka, A., Vassar, M.J., Vespa, P., Wang, K., Yue, J.K., and Zafonte, R. (2019). Recovery after mild traumatic brain injury in patients presenting to US Level I Trauma Centers: a Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. JAMA Neurol. 76, 1049–1059. Erratum in: JAMA Neurol. doi: 10.1001/jamaneurol.2019.3698.
26. Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
27. University of Michigan; SIREN Network. (2018). BOOST-3; Brain Oxygen Optimization in Severe TBI Phase-3. siren.network/clinical-trials\boost-3 (Last accessed April 12, 2021).