



Published in final edited form as:

*J Head Trauma Rehabil.* 2010 ; 25(5): 375–382. doi:10.1097/HTR.0b013e3181d27fe3.

## Measuring Outcome in Traumatic Brain Injury Treatment Trials: Recommendations From the Traumatic Brain Injury Clinical Trials Network

**Dr. Emilia Bagiella, PhD, Dr. Thomas A. Novack, PhD, Dr. Beth Ansel, PhD, Dr. Ramon Diaz-Arrastia, MD, Dr. Sureyya Dikmen, PhD, Dr. Tessa Hart, PhD, and Dr. Nancy Temkin, PhD**  
Department of Biostatistics, Mailman School of Public Health, Columbia University, New York (Dr Bagiella); Physical Medicine and Rehabilitation, University of Alabama at Birmingham (Dr Novack); National Center for Medical Rehabilitation Research, National Institute of Child Health and Human Development, Bethesda, Maryland (Dr Ansel); Department of Neurology, University of Texas, Southwestern Medical Center, Dallas (Dr Diaz-Arrastia); Departments of Rehabilitation Medicine, Division of Clinical Psychology and Neuropsychology, School of Medicine (Dr Dikmen) and Neurological Surgery and Biostatistics (Dr Temkin), University of Washington, Seattle; and Moss Rehabilitation Research Institute, Philadelphia, Pennsylvania (Dr Hart)

### Abstract

**Background**—Traumatic brain injury (TBI) involves several aspects of a patient’s condition, including physical, mental, emotional, cognitive, social, and functional changes. Therefore, a clinical trial with individuals with TBI should consider outcome measures that reflect their global status.

**Methods**—We present the work of the National Institute of Child Health and Development–sponsored Traumatic Brain Injury Clinical Trials Network Outcome Measures subcommittee and its choice of outcome measures for a phase III clinical trial of patients with complicated mild to severe TBI.

**Results**—On the basis of theoretical and practical considerations, the subcommittee recommended the adoption of a core of 9 measures that cover 2 different areas of recovery: functional and cognitive. These measures are the Extended Glasgow Outcome Scale; the Controlled Oral Word Association Test; the Trail Making Test, Parts A and B; the California Verbal Learning Test–II; the Wechsler Adult Intelligence Scale–III Digit Span subtest; the Wechsler Adult Intelligence Scale–III Processing Speed Index; and the Stroop Color-Word Matching Test, Parts 1 and 2.

**Conclusions**—The statistical methods proposed to analyze these measures using a global test procedure, along with research and methodological and regulatory issues involved with the use of multiple outcomes in a clinical trial, are discussed.

### Keywords

clinical trial; global test procedure; outcome measures; traumatic brain injury

---

Traumatic brain injury (TBI) is a public health problem of major proportions. Each year, more than 230 000 people in the United States alone sustain a TBI, resulting in hospitalization and

---

Corresponding Author: Emilia Bagiella, PhD, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 W 168th St, New York, NY 10032 (eb51@columbia.edu).

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal’s Web site ([www.headtraumarehab.co](http://www.headtraumarehab.co)).

potential life-long disability due to a complex variety of cognitive, physical, and emotional sequelae.<sup>1</sup> Despite the social and economic impact of TBI, few treatments have been shown conclusively to ameliorate the adverse outcomes. In recent years, there have been calls for more definitive trials, both to prevent the effects of TBI and to remedy its long-term effects.<sup>2</sup> Such trials must grapple with multiple challenges posed by the diverse effects of TBI along the continuum of severity, as well as the practical constraints of data collection. The purpose of this article is to discuss the scientific and pragmatic decisions made by a group of clinical researchers faced with the crucial decision of selecting outcome measures to assess treatment efficacy in patients with TBI.

Choosing the appropriate primary-outcome measures for a clinical trial is a critical step to minimize the potential of type I or II errors. Such errors, which refer, respectively, to over- or underestimating the true effects of treatment, can have a profound effect on whether an intervention is employed in the clinical care of patients and on future trials involving the same patient group. In clinical trials, the primary outcome measure is the one that is used to answer the study's core question. Typically, the primary outcome is a parameter of important clinical interest that best represents the mechanism of action of the intervention of interest or a variable on which the treatment may have the greatest effect. For example, a study of a treatment focused on improving basic mechanisms of attention after TBI might use a psychometric measure of sustained attention or information-processing speed as a primary outcome. In contrast, a study testing the effects of a multifaceted treatment program directed to return to work after TBI might use return-to-work rates, number of hours successfully worked, or employer or employee satisfaction ratings as a primary outcome. From a methodological point of view, the choice of the primary outcome measure is important because it affects the design, sample-size calculation, and data-analysis plan of the study.

In designing a clinical trial, usually a single primary outcome and a number of secondary outcomes are selected. To continue with the examples mentioned earlier, secondary outcomes of the attention treatment might be degree of engagement in therapy as judged by therapists and ratings of attentiveness in the home setting provided by family members. A secondary outcome of the vocational treatment might be improved self-ratings on a scale measuring quality of life. Selecting a single, primary outcome makes the trial design cleaner and simpler and helps avoid inconsistent results and inappropriate conclusions. The sample size calculation, statistical stopping rules, and major conclusions about treatment efficacy are generally based on the primary outcome, while secondary outcomes are used to support evidence of treatment efficacy, confirm consistency of the results, strengthen the internal validity of the study, and generate new hypotheses.

In early phase studies, the specific mechanism of action that is investigated (mechanistic studies), a particular biological parameter of interest (proof-of-concept studies), or the need to detect a first sign of treatment efficacy (nonsuperiority or multistage phase II studies) mandates the selection of a single primary outcome. In more advanced stages of development of the drug or other treatment, however, like phase III trials, selecting a single, primary outcome measure, although convenient methodologically, is difficult. Complexities arise especially for those clinical conditions with widespread effects at different levels of function and for treatments that may affect multiple types of outcomes. A common concern among both clinicians and statisticians is the risk of increasing the type II error that can occur when a promising treatment is discarded because it is found ineffective on a primary outcome measure, while effective on other measures. Researchers in many fields have considered using several primary outcomes to represent the complexity and multiplicity of patients' responses to interventions (eg, stroke).<sup>3</sup> In this article, we discuss the choice of outcome measures for a phase III clinical trial of a pharmacologic intervention involving patients with complicated mild to severe TBI.

## MATERIALS AND METHODS

The Traumatic Brain Injury Clinical Trials (TBI-CT) Network is a consortium comprising 8 clinical sites and 1 data-coordinating center, funded by the National Institute of Child Health and Development to conduct phase III multicenter clinical trials to improve the outcomes of people experiencing TBI. In contrast to multi-center networks that were established primarily for specialized clinical care, such as the centers in the Defense and Veterans-Brain Injury Center program, the TBI-CT network was created to implement rigorous clinical trials within the acute or rehabilitation phases at high-volume centers of clinical excellence for TBI. This focused purpose stands also in contrast to the TBI Model Systems program, which supports a wide variety of studies on TBI but focuses more on longitudinal research than clinical trials. The first trial designed within the network is a multicenter, randomized, double-blind, placebo-controlled study of the effect of citicoline in patients with TBI (COBRIT). The trial is designed to enroll 1426 patients with severe, moderate, and complicated mild TBI (Glasgow Coma Scale score, 3–15). Details about the design of this trial and baseline characteristics of the sample are in a separate article published.<sup>4</sup>

One of the first challenges presented to the TBI-CT Network was to identify and select a set of outcome measures that would be used as primary outcomes for COBRIT. To this end, a subgroup of Network investigators, the Outcome Measures Subcommittee, including neurosurgeons, neurologists, psychiatrists, neuropsychologists, and statisticians, met regularly for more than a year, with the goal to review a set of outcome measures suitable for the TBI population and to choose measures to be used as primary outcomes in TBI-CT Network trials.

The chief concern of the Outcome Measure Subcommittee was to identify measures that together would reflect the “global” status of patients with TBI. Traumatic brain injury has diffuse cerebral effects, which in turn cause an array of impairments and disabilities in functional, physical, emotional, cognitive, and social spheres. It was felt that no single measure could capture the multidimensional nature of the outcome of TBI. Thus, the Subcommittee was mandated to identify measures that would reflect areas likely to be affected by TBI and that would document improvement during recovery. It was assumed that multiple measures would be necessary to address the breadth of potential deficits and recovery following TBI, and to capture important outcomes at levels of function from impairment to societal participation.

Not only the multifaceted effects of TBI but also the timing of outcome assessment presents particular problems in research and clinical care of patients with this disability. The primary outcome measure for an acute or emergent intervention (often survival) will have little meaning a year after injury when the focus may be on resumption of independent living. Because TBI often occurs in a relatively young population, the assessment of outcome presents several unique challenges that are not often encountered in other clinical conditions such as cancer or cardiac disease. In TBI, relevant outcomes may vary throughout the course of recovery and may be measured several years after the event. Therefore, it seems unreasonable to expect a single outcome measure to capture the extent of recovery over time following TBI. In addition, assessing outcome on certain measures may not be feasible for some study participants. Assessment of cognitive or behavioral functioning is not possible for a person in a minimally conscious state, for instance. In many past studies, with a few exceptions,<sup>5–9</sup> the inability to assess outcome because of cognitive impairment has resulted in the cognitive data being considered missing. The Subcommittee felt strongly that the inability of some participants to take part in outcome assessment due to the severity of the neurological disorder needed to be incorporated into the data in a meaningful way. Classification of such cases as missing data may bias the treatment comparison and exclude the outcome of the most impaired patients from consideration.

A third issue the Outcome Measure Subcommittee had to address was that outcome of TBI is influenced by many factors. These factors include the severity and characteristics of the brain injury, injuries sustained to other parts of the body in the same incident, preinjury functional, physical, cognitive, and demographic characteristics of the person injured, and time from the injury to when the outcome is assessed. The challenge, therefore, for researchers in TBI is to develop and select efficient and valid outcome measures that are able to cover the full spectrum of TBI severity and capture important constructs of recovery.

Bearing all of these considerations in mind, a primary task of the Outcome Measure Subcommittee was to select measures that would prove sensitive to the effects of the treatment, citicoline, throughout the spectrum of injury severity. Citicoline is a naturally occurring compound that may have general neuroprotective effects via such mechanisms as enhancement of cerebral blood flow.<sup>10</sup> However, careful measurement of new learning and memory, and other cognitive functions, was mandated by the anticipated effects of citicoline specifically on the cholinergic system implicated in learning and memory.<sup>11,12</sup>

COBRIT is designed to enroll patients as close as possible to the time of injury with the aim of improving the underlying pathology and associated functional and cognitive functions with ultimate benefit to global outcome. Thus, the Subcommittee recommended that the primary outcome focus on measures of global functional status *and* cognitive abilities. Cognitive status and global outcome are typically correlated in TBI, presumably because both are affected strongly by changes in the neurologic substrate. In contrast, outcomes such as emotional status and quality of life are important in their own right but are also more likely to be affected by factors external to the neural substrate (eg, premorbid personality, family status, environmental factors). Thus, the latter types of outcomes were designated for COBRIT as secondary.

## RESULTS

On the basis of the literature review<sup>13–17</sup> and results of ongoing trials, the Outcome Measures Subcommittee analyzed and discussed 35 possible measures. Based on the above-mentioned considerations, 9 of these measures were chosen as the Core Outcome Measures for COBRIT. For a global functional measure, the Extended Glasgow Outcome Scale (GOS-E)<sup>18</sup> was chosen. For neuropsychological measures, the Controlled Oral Word Association Test,<sup>19</sup> the Trail Making Test,<sup>20,21</sup> and the Stroop Color-Word Matching Test<sup>22</sup> were chosen as measures of executive functions. The California Verbal Learning Test–2 was selected as a measure of episodic memory.<sup>23</sup> The Digit Span Test (from the Wechsler Adult Intelligence Scale–III<sup>24</sup>) was chosen to assess span of attention/working memory and the Processing Speed Index (also from the Wechsler Adult Intelligence Scale–III<sup>24</sup>) was selected as a measure of information-processing speed.

These measures were chosen because they assess major constructs that are known to be affected by TBI and were judged likely to be sensitive to cholinergic treatment. They have good psychometric properties, cover a broad range of functioning, and are commonly used clinical neuropsychological measures accepted as standards in the field. Furthermore, these measures or other measures examining the same constructs have previously been used in TBI studies.<sup>8, 9,25–27</sup> Moreover, the measures of cognitive impairment correlate well with both measures of disability, like the GOS-E, and other indicators of functional status.<sup>28–32</sup>

Pragmatic considerations also affected the selection of outcome measures for COBRIT. Outcome assessment should not represent an undue burden for the study participants or the study personnel. It was expected that under normal conditions the entire battery of measures could be administered in about 1 hour. The order of test administration was fixed so that any

effects of fatigue would be consistent across subjects and more [www.headtraumarehab.com](http://www.headtraumarehab.com) challenging tests such as the California Verbal Learning Test-2 would come early in the order.

### Test completion codes

It was anticipated that the ability of study participants to complete the outcome battery would vary, particularly for patients with severe injury who were early in the course of recovery. Therefore, the group believed that it was important to capture the reasons for missing values in order to distinguish the inability to provide data for reasons related to the TBI from unrelated reasons such as unknown whereabouts of the participant, unwillingness to provide data, etc. To account for these distinctions, test completion codes were developed to identify reasons for a person's inability to complete a test (Table 1). The first code identified a score that is considered valid; that is, the measure was administered correctly and the participant completed it to the best of his or her ability. The second code was intended for situations in which the participant is unable to complete the test for neurological reasons related directly or indirectly to TBI. A third code was created to identify situations in which the test was not completed for non neurological reasons, such as injury other than to the brain, obvious lack of effort or cooperation, problems with English proficiency, or interruption of the testing session. Fourth and fifth codes were added to reflect site-specific reasons for missing scores.

The distinctions made with these codes are very important to the research endeavor. On tests of ability such as the assessments of memory, attention, and executive function, participants unable to complete the tests because of brain-related impairment can be categorized as exhibiting a very poor outcome, for example, by assigning the worst possible score, or a score worse than those observed. If the outcome is expressed as a dichotomous variable (poor response vs good response) those patients unable to complete the test because of neurological reasons can be easily classified as exhibiting unfavorable responses. Therefore, no informative data are lost for analysis purposes. Test completion codes of 3, 4, or 5 are classified as missing data. Missing data on the tests that do not measure cognitive ability, for example, the emotional function measure, must be left as missing because there is no meaningful "lowest score" that captures the effects of severe TBI.

### Statistical and methodological issues

From a methodological and statistical point of view, the use of multiple primary outcome measures raises a number of issues. First, if a single measure is used to calculate the study sample size and power, it may be difficult to determine the appropriate measure to use. Choosing a measure that is likely to change as a result of the intervention may lead to a study that is underpowered to test important hypotheses involving the other measures. Choosing the least promising measure, on the other hand, may lead a study that is not feasible because of the large sample required. Second, the choice of several outcome measures inevitably raises the issue of multiple comparisons. While it would be very interesting to determine the effect of the treatment on each individual outcome measure by performing several statistical tests, such a procedure would require controlling the experiment-wise error rate to avoid an increase in false-positive findings. Current procedures used to maintain the experiment-wise error rate at the nominal level (say 0.05), for example, the Bonferroni procedure or similar approaches, may become excessively conservative and inefficient in the context of multiple outcome end points. It is important to recognize that these procedures are based on the assumption that the multiple tests are independent. This is often not the case with multiple outcome measures, which are generally moderately correlated and are expected to behave qualitatively similarly. Moreover, as the number of tests increases, the correction becomes overly conservative and rejection of the null hypothesis for any of the measures becomes unlikely, with a potential type II error forthcoming. This can occur even in a situation when all measures show a moderate effect of an intervention but none show a very strong effect.



Multivariate methods, such as the Hotelling  $T^2$  test or multivariate analysis of variance, are also inefficient in the context of multiple outcome measures in clinical trials. These tests are nonspecific because they are designed to detect any departure from the null hypothesis rather than specific alternatives that are biologically or clinically meaningful. Moreover, these methodologies require a large sample size to detect small differences in selected variables.

**A global test procedure**—A global test procedure represents an efficient approach for sample size calculation and statistical analysis of studies with several outcome measures. The procedure is based on the assumption that the treatment effect is constant across all measures. Once a common effect size is established, the global test procedure tests the null hypothesis that the common effect size is zero against an alternative hypothesis that the common effect size is different from zero.

The global test procedure offers several advantages compared with the multivariate procedures or the multiple testing correction procedures. It offers a method to utilize several outcome measures without the need to prespecify one as primary; it avoids loss of power due to multiple comparisons as it tests all measures at once at the nominal statistical level; it has greater statistical power than any single outcome measure because it uses all the available information from all measures; it is based on sound statistical theory; and it is easily interpreted.

For COBRIT, the Outcome Measure Subcommittee decided that an approach based on a binary outcome (favorable response vs unfavorable response) would be used. Dichotomizing the outcome measures to reflect success or failure simplifies the problem of assigning a score to a patient too cognitively impaired to be tested and it facilitates dealing with missing data. If a patient fails to complete the core battery, imputing the dichotomous score for the incomplete scales using the response to the completed scales is straightforward. Moreover, reporting of the study results in terms of treatment success or failure can be more easily interpreted for clinical significance. While it is true that dichotomization may be arbitrary, it avoids the similar arbitrary scaling issues with quantitative outcome measures. To examine more nuanced relationships between the treatment and particular outcome measures, outcome scores in their original units, whether raw or standardized, can always be evaluated in secondary analysis.

To construct the global test procedure the GOS-E was dichotomized as 1 to 6 for unsuccessful outcome and 7 to 8 for successful outcome. All other measures in [www.headtraumarehab.com](http://www.headtraumarehab.com) the battery are continuous. The raw scores were used to categorize participants in the successful outcome group (if the score is no worse than 1 standard deviation from the mean of the normative population for that scale) or in the unsuccessful outcome group (if the raw score is beyond 1 standard deviation below the mean for the normative population for that scale).

The normative values used to determine the cutoffs in this context are very important. As a general rule, it is best to use values established for unimpaired individuals similar to the clinical population to be studied. Census-based norms may be suitable in some studies, but not in others. For COBRIT, values were available for most tests in the battery for trauma cases (without head injury) obtained at the University of Washington.<sup>37</sup>

**An example**—The global test procedure for the COBRIT study is based on a logistic model that provides an estimate of the common effect size expressed as a global odds ratio (OR) for successful outcome in the treatment group compared with the placebo group, along with a  $P$  value,  $z$  score, standard error, and confidence interval. The mathematical details of the global test for binary outcomes are presented in an Appendix (see Supplemental Digital Content, details of the global test procedure for binary outcome, see Appendix, SDC Content 1, <http://links.lww.com/JHTR/A44>).

Data from previous trials were used to define the common effect size as an OR = 1.4. This corresponds to an absolute improvement, under the experimental condition, of about 8% on the GOS-E and of 13% to 20% on the cognitive measures. Fixing the type I error at 0.05 and the power at 85% and assuming a moderate to high correlation among the measures (0.25–0.81), it was calculated that the trial needed 1240 patients to detect the effect size of interest. This sample size was then adjusted to reach 1426 patients to account for attrition. Had the Network decided to base the sample size calculation on a single primary outcome (eg, the GOS-E), 1836 participants would have been required to run a trial with the same power, type I error and effect size. Thus, using the global test procedure resulted in 32% fewer study participants being required.

**Testing the treatment effect on the individual measures**—Rejecting the global null hypothesis (ie, the hypothesis that none of the outcome measures shows a treatment effect) may not be sufficient to determine the specific measures on which the treatment has an effect. To investigate this and at the same time avoid a multiple comparison artifact, a closed test procedure can be used<sup>43</sup> (see Supplemental Digit Content, procedure for testing the effects of treatment on specific measures). The procedure is based on a stepwise analysis in which, after rejection of the null global hypothesis, all possible subsets of hypotheses can be tested in a hierarchical fashion. The procedure also allows testing single hypotheses as well as groups of hypotheses, thus allowing for detection of possible patterns of treatment effect.

## DISCUSSION

Traumatic brain injury often results in a multifaceted disability, and recovery from TBI is a complex process that involves many aspects of functional and cognitive change. Therefore, measuring outcome presents many challenges in large phase III trials of patients with TBI. Targeting only 1 aspect of possible improvement may not be sufficient to determine the effectiveness of a new intervention and it is susceptible to an increased type II error. To address this issue, the TBI-CT Network Outcome Subcommittee has proposed to use a core of 9 measures that in combination will evaluate outcome in the Network's first trial. These measures cover 2 components of important outcome: functional status and cognitive abilities.

Many clinical trials of patients with TBI still rely on the GOS<sup>33</sup> as the primary measure of recovery and to determine treatment effect. In a recent article, Lu et al<sup>38</sup> point out the pitfalls associated with the use of the GOS in TBI trials and, in particular, the vulnerability of this scale to misclassification. The authors illustrate how common types of misclassification result in considerable loss of statistical power and the attenuation of the true treatment effect. The same authors also suggest that the extended scale, the GOS-E, suffers from the same misclassification bias.

Other authors<sup>44</sup> have suggested a dichotomous outcome over the 5 GOS categories in an attempt to maximize the sensitivity of the scale. Murray and colleagues<sup>45</sup> described a statistical methodology, the sliding dichotomy, that may improve the statistical power of the GOS. The sliding dichotomy is a statistical model-based procedure that uses information on patient's baseline characteristics to predict that patient's outcome. On the basis of this procedure, a favorable outcome is defined as better than would be expected, taking account of each individual patient's baseline prognosis. The main advantage of this procedure is that response is tailored to individual patients rather than being fixed equal for all study participants. Therefore, patients with poor prognostic characteristics who are expected to make less progress will have a lower cutoff than patients with good prognostic variables. Compared with a procedure in which the cutoff is fixed and constant across severity groups and other prognostic factors, the sliding dichotomy procedure increases the chance that even severe patients reach a favorable outcome, thus increasing the power of the study.

Contrary to this approach, the Outcome Measures Subcommittee chose a fixed cutoff because it decided that it was important to determine whether an intervention works according to very general standards; for example, does the intervention increase the proportion of patients who are able to return to their previous life activities (eg, back to family and work). This type of outcome has a wider public health as well as social importance. An intervention that improves outcome from vegetative to severe disability status in a subset of the patient population, although of great importance to the individual, will not have the same impact in a more general setting. Moreover, the interpretation of favorable response according to the sliding dichotomy is not straightforward and not generalizable whereas the use of a fixed cutoff makes it easier to translate to a different or a more general population.

Another difficulty of the sliding dichotomy approach is that the methodology is model-based and thus requires the existence of an existing data set to build the model and determine the patient-specific cutoffs. Because it is not known whether the cutoffs determined in other studies, with a specific set of covariates, also apply to a newly designed study with possibly a different set of covariates, the only way to use this procedure is to wait until the end of the study and then use the collected data to construct a model and determine the cutoffs. This approach is obviously data driven and is subject to criticism as being post hoc. On the other hand, fixed cutoffs can be declared a priori and do not depend on any specific model.

When specifying fixed cutoffs some concern may arise regarding ceiling or floor effects that may result by including very impaired or very intact patients in the sample. If a large proportion of patients are too impaired to improve or are already intact enough to score above the cutoff at baseline, the power of the study may decrease. For the COBRIT trial, the Outcome Subcommittee considered that this would not be an issue since the trial does not cover the whole spectrum of TBI but rather focuses on a sample of patients who, according to published research, have injury to the brain substantial enough to avoid ceiling effects on the selected measures.

The approach proposed by the TBI-CT Network Outcome Subcommittee uses a dichotomized version of the GOS-E. However, because the outcome is measured by 8 additional scales the risk of loss of power and attenuation of the effect size due to misclassification is greatly minimized as compared with a study that relies on the GOS-E alone.

Another concern about the global test procedure pertains to the assumption of equal effect size across all measures. Although the global test procedure can be carried out even if not all outcomes have, in fact, the same effect size, power may be reduced if some outcomes depart significantly from the common effect size. However, if that happened, it would indicate an ambiguous situation, in which a treatment would be beneficial on some outcomes and inert or even harmful on others, a situation in which careful judgment would be required. Input from independent, objective sources may also prove useful in the evaluation of ambiguous results.

A final consideration concerns the use of the multiple outcome measures in relation to regulatory agencies. Regulatory approval in the form of an investigational new drug or a new drug approval by the Food and Drug Administration (FDA) is often required before a new drug is used in a clinical trial. The FDA has very clear guidelines about the design, sample size calculation, statistical analysis, and choice of the primary outcome. For trials addressing rehabilitation and treatment of TBI, the FDA has historically recommended the use of the GOS-E as the single primary outcome. The use of a global test procedure to simultaneously test several outcomes was proposed to the FDA and granted approval in the past. The TBI-CT Network was also successful in securing approval from the FDA to use the binary global test procedure for the Network's first trial.



Measuring outcome in trials of patients with TBI presents several challenges. We believe that our choice of a global outcome approach for the COBRIT trial addresses many of these issues and will provide a robust and comprehensive estimate of the treatment effect.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The work of the Traumatic Brain Injury Clinical Trials Network Outcome Measures Subcommittee is supported by National Institute of Child Health and Development grants U01HD042823, U01HD042738, U01HD042687, U01HD042736, U01HD042678, U01HD042686, U01HD042653, U01HD042689, and U01HD042652 to 8 clinical sites and a data-coordinating center.

## References

1. Thurman DJ, Alverson C, Dunn KA, Guerrero J, Sniezek JE. Traumatic brain injury in the United States: a public health perspective. *J Head Trauma Rehabil* 1999;14(6):602–615. [PubMed: 10671706]
2. Ragnarsson KT. Traumatic brain injury research since the 1998 NIH Consensus Conference: accomplishments and unmet goals. *J Head Trauma Rehabil* 2006;21(5):379–387. [PubMed: 16983223]
3. Tilley BC, Marler J, Geller NL, et al. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial. *Stroke* 1996;27(11):2136–2142. [PubMed: 8898828]
4. Zafonte R, Friedewald WT, Lee SM, et al. Citicoline brain injury treatment (COBRIT) trial: design and methods. *J Neurotrauma* 2009;26:2207–2216. [PubMed: 19803786]
5. Temkin NR, Anderson GD, Winn HR, et al. Magnesium sulfate for neuroprotection after traumatic brain injury: a randomised controlled trial. *Lancet Neurol* 2007;6(1):29–38. [PubMed: 17166799]
6. Temkin NR, Dikmen SS, Anderson GD, et al. Valproate therapy for prevention of posttraumatic seizures: a randomized trial. *J Neurosurg* 1999;91(4):593–600. [PubMed: 10507380]
7. Temkin NR, Dikmen SS, Wilensky AJ, Keihm J, Chabal S, Winn HR. A randomized, double-blind study of phenytoin for the prevention of post-traumatic seizures. *N Engl J Med* 1990;323(8):497–502. [PubMed: 2115976]
8. Dikmen SS, Machamer JE, Winn HR, Anderson GD, Temkin NR. Neuropsychological effects of valproate in traumatic brain injury: a randomized trial. *Neurology* 2000;54(4):895–902. [PubMed: 10690983]
9. Dikmen SS, Temkin NR, Miller B, Machamer J, Winn HR. Neurobehavioral effects of phenytoin prophylaxis of posttraumatic seizures. *JAMA* 1991;265(10):1271–1277. [PubMed: 1995974]
10. Leon-Carrion J, Dominguez-Roldan JM, Murillo-Cabezas F, del Rosario Dominguez-Morales M, Munoz-Sanchez MA. The role of citicholine in neuropsychological training after traumatic brain injury. *NeuroRehabilitation* 2000;14(1):33–40. [PubMed: 11455066]
11. Fioravanti M, Yanagi M. Cytidinediphosphocholine (CDP-choline) for cognitive and behavioural disturbances associated with chronic cerebral disorders in the elderly. *Cochrane Database Syst Rev* 2005;(2):CD000269. [PubMed: 15846601]
12. Levin HS. Treatment of postconcussional symptoms with CDP-choline. *J Neurol Sci* 1991;103(suppl):S39–S42. [PubMed: 1940965]
13. COMBI. The Center for Outcome Measurement in Brain Injury. [Accessed 2006]. <http://www.tbims.org/combi/bg.html>
14. Clifton GL, Kreutzer JS, Choi SC, et al. Relationship between Glasgow Outcome Scale and neuropsychological measures after brain injury. *Neurosurgery* 1993;33(1):34–39. [PubMed: 8355845]
15. Jennet, B. *Head Injury*. London: Chapman & Hall; 1997. Outcome after severe head injury; p. 439-461.

16. Wilson JT, Pettigrew LE, Teasdale GM. Emotional and cognitive consequences of head injury in relation to the Glasgow Outcome Scale. *J Neurol Neurosurg Psychiatry* 2001;70:267–268.
17. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: guidelines for their use. *J Neurotrauma* 1998;15(8):573–585. [PubMed: 9726257]
18. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: guidelines for their use. *J Neurotrauma* 1998;15(8):573–585. [PubMed: 9726257]
19. Benton, AL.; Hamsher, KD.; Sivan, AB. Multilingual Aphasia Examination. Manual of Instructions. 3. Iowa City, IA: AJA Associates; 2000.
20. Reitan, RM. Trail Making Test. Manual for Administration and Scoring. Tucson, AZ: Reitan Neuropsychology Laboratory; 1992.
21. Reitan, RM.; Wolfson, D. The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. 2. Tucson, AZ: Neuropsychology Press; 1993.
22. Dodrill CB. A neuropsychological battery for epilepsy. *Epilepsia* 1978;19(6):611–623. [PubMed: 738230]
23. Delis, DC.; Kramer, JH.; Edith, K.; Ober Beth, A. California Verbal Learning Test. 2. San Antonio, TX: PsychCorp; 2000. Adult Version
24. Wechsler, D. Wechsler Adult Intelligence Scale (WAIS-III). 3. San Antonio, TX: The Psychological Corporation; 1997.
25. Kreutzer JS, Gordon WA, Rosenthal M, Marwitz J. Neuropsychological characteristics of patients with brain injury: preliminary findings from a multicenter investigation. *J Head Trauma Rehabil* 1993;8(2):47–59.
26. Millis SR, Rosenthal M, Novack TA, et al. Long-term neuropsychological outcome after traumatic brain injury. *J Head Trauma Rehabil* 2001;16(4):343–355. [PubMed: 11461657]
27. Vanderploeg RD, Crowell TA, Curtiss G. Verbal learning and memory deficits in traumatic brain injury: encoding, consolidation, and retrieval. *J Clin Exp Neuropsychol* 2001;23(2):185–195. [PubMed: 11309672]
28. Doctor JN, Castro J, Temkin NR, Fraser RT, Machamer JE, Dikmen SS. Workers' risk of unemployment after traumatic brain injury: a normed comparison. *J Int Neuropsychol Soc* 2005;11(6):747–752. [PubMed: 16248910]
29. Ross SR, Millis SR, Rosenthal M. Neuropsychological prediction of psychosocial outcome after traumatic brain injury. *Appl Neuropsychol* 1997;4:165–170. [PubMed: 16318480]
30. Dikmen S, Temkin N, Machamer J, Holubkov A, Fraser R, Winn HR. Employment following traumatic head injuries. *Arch Neurol* 1994;51(2):177–186. [PubMed: 8304843]
31. Sherer M, Novack TA, Sanderm AM, Struchenm MA, Aldersonm A, Nakase-Thompson R. Neuropsychological assessment and employment outcome after traumatic brain injury: a review. *Clin Neuropsychol* 2002;16:157–178. [PubMed: 12221479]
32. Novack TA, Bush BA, Meythaler JM, Canupp K. Outcome after traumatic brain injury: pathway analysis of contributions from premorbid, injury severity, and recovery variables. *Arch Phys Med Rehabil* 2001;82(3):300–305. [PubMed: 11245749]
33. Jennett B, Bond M. Assessment of outcome after severe brain damage. A practical scale. *Lancet* 1975;1:480–484. [PubMed: 46957]
34. Levin HS, Boake C, Song J, et al. Validity and sensitivity to change of the Extended Glasgow Outcome Scale in mild to moderate traumatic brain injury. *J Neurotrauma* 2001;18(6):575–584. [PubMed: 11437080]
35. Benton, AL.; Sivan, AB.; Hamsher, KD.; Varney, NR.; Spreen, O. Contributions to Neuropsychological Assessment: A Clinical Manual. New York, NY: Oxford University Press; 1994.
36. Reitan RM, Wolfson D. The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. *Arch Clin Neuropsychol* 2004;19(2):281–288. [PubMed: 15010091]
37. Dikmen S, Machamer JE, Winn HR, Temkin N. Neuropsychological outcome at 1 year post head injury. *Neuropsychology* 1995;9(1):80–90.

38. Lu J, Murray GD, Steyerberg EW, et al. Effects of Glasgow Outcome Scale misclassification on traumatic brain injury clinical trials. *J Neurotrauma* 2008;25:641–651. [PubMed: 18578634]
39. Hart T, Millis S, Novack TA, Englander J, Fidler-Shepherd R, Bell K. The relationship between neuropsychological function and level of caregiver supervision at 1 year post traumatic brain injury. *Arch Phys Med Rehabil* 2003;84:221–230. [PubMed: 12601653]
40. Rapport LJ, Hanks RA, Millis SR, Deshpande SA. Executive functioning and predictors of falls in the rehabilitation setting. *Arch Phys Med Rehabil* 1998;79(6):629–633. [PubMed: 9630140]
41. Wechsler, DW. *Adult Intelligence Scale (WAIS-III)*. San Antonio, TX: The Psychological Corporation; 1997.
42. Kersel DA, Marsh NV, Havill JH, Sleigh JW. Neuropsychological functioning during the year following severe traumatic brain injury. *Brain Inj* 2001;15(4):283–296. [PubMed: 11299130]
43. Lehmacher W, Wasserman G, Reitmer P. Procedures for two sample comparisons with multiple endpoints controlling for the experimentwise error rate. *Biometrics* 1991;47:511–521. [PubMed: 1912258]
44. Choi SC, Clifton GL, Marmarou A, Miller ER. Misclassification and treatment effect on primary outcome measures in clinical trials of severe neurotrauma. *J Neurotrauma* 2002;19(1):17–22. [PubMed: 11852975]
45. Murray, DM. *Design and Analysis of Group-Randomized Trials*. New York, NY: Oxford University Press; 1998.

TABLE 1

Description of the Traumatic Brain Injury Clinical Trials network core outcome measures

Measure	Description	Designed to measure	Properties
GOS-E (Glasgow Outcome Scale Extended) <sup>17</sup>	Administered by structured interview with a focus on independence in the home and community as changed by traumatic brain injury (TBI). Results classify the outcome at 8 levels ranging from death to upper-good recovery	Global outcome; developed to improve sensitivity and reliability over the original GOS <sup>33</sup>	Relationship with initial severity and multiple measures of TBI outcome superior to original GOS. More sensitive to change from 3 to 6 mo than GOS <sup>34</sup>
COWAT (Controlled Oral Word Association Test) <sup>35</sup>	Subject generates as many words as possible beginning with a given letter within 1 min. Score is the total number of words generated over 3 trials	Verbal fluency under time pressure, thought to reflect frontal/executive function	Sensitive to effects of TBI <sup>25,34</sup>
Trail Making Test <sup>36</sup>	Two components, A and B, involve connecting items sequentially in a visual array. Part A requires simple number sequencing while part B requires alternation between numbers and letters. Scores are the times needed for completion	Sustained and divided attention; mental flexibility; psychomotor speed, set-shifting ability	Sensitive to TBI severity, to recovery over time and related to other measures of outcome. <sup>37,38</sup> Correlated with functional outcomes such as need for supervision <sup>39</sup> and functional safety <sup>40</sup>
California Verbal Learning Test <sup>23</sup>	Learning of a 16-word list over 5 trials. Scores include total number of words recalled, number within categories, recognition memory, etc	Verbal learning, use of category information to promote recall	Sensitive to the effects of TBI and its severity of injury <sup>27,34</sup>
Digit span (from WAIS III) <sup>41</sup>	Maximum span for repeating a series of digits forward and backward	Auditory working memory	Sensitive to severity of injury and recovery <sup>42</sup>
Processing speed index (from Wechsler Adult Intelligence Scale-III) <sup>41</sup>	Two components: Digit-symbol coding: Rapid transcription of symbols Symbol search: Rapid visual scanning of geometric forms	Visual processing speed; visuomotor speed	Has a component of working memory; sensitive to the effects of TBI <sup>35</sup>
Stroop Task <sup>22</sup>	Generates a conflict between automatic and effortful	Executive abilities, particularly inhibition of automatic responses	Sensitive to effects of TBI even in those with good recovery <sup>39</sup>

Measure	Description	Designed to measure	Properties
	processing by requiring identification of color of ink ignoring printed words (eg, "red" written in green ink)		