Original Articles

Impact of GOS Misclassification on Ordinal Outcome Analysis of Traumatic Brain Injury Clinical Trials

Juan Lu,[1,2] Anthony Marmarou,[1] and Kate L. Lapane,[1] on behalf of the IMPACT investigators

Abstract

This study extends our previous investigation regarding the effect of nondifferential dichotomous Glasgow Outcome Scale (GOS) misclassification in traumatic brain injury (TBI) clinical trials to the effect of GOS misclassification on ordinal analysis in TBI clinical trials. The impact of GOS misclassification and ordinal outcome analysis was explored via probabilistic sensitivity analyses using TBI patient datasets from the IMPACT database ($n = 9205$). Three patterns of misclassification were explored given the pre-specified misclassification distributions. For the random pattern, we specified a trapezoidal distribution (minimum: 80%, mode: 85%, and 95%, maximum: 100%) for both sensitivity and specificity; for the upward pattern, the same trapezoidal distribution for sensitivity but with a perfect specificity; and for the downward pattern, the same trapezoidal distribution for specificity but with a perfect sensitivity. The conventional 95% confidence intervals and simulation intervals, which accounts for the misclassification and random errors together, were reported. The results showed that given the specified misclassification distributions, the misclassification with a random or upward pattern would have caused a slightly underestimated outcome in the observed data. However, the misclassification with a downward pattern would have resulted in an inflated estimation. Thus the sensitivity analysis suggests that the nondifferential misclassification can cause uncertainties on the primary outcome estimation in TBI trials. However, such an effect is likely to be small when ordinal analysis is applied, compared with the impact of dichotomous GOS misclassifications. The result underlines that the ordinal GOS analysis may gain from both statistical efficiency, as suggested by several recent studies, and a relatively smaller impact from misclassification as compared with conventional binary GOS analysis.

Key words: clinical trials; Glasgow Outcome Scale; misclassification; probability sensitivity analysis

Introduction

SEVERAL RECENT STUDIES have explored the ordinal analysis of the Glasgow Outcome Scale (GOS; Bath et al., 2008; Jennett and Bond, 1975; McHugh et al., 2010; Optimizing Analysis of Stroke Trials [OAST] Collaboration, 2007), and other ordinal outcomes commonly used in traumatic brain injury (TBI) and stroke clinical trials. Previously, studies have dichotomized GOS outcomes that split the GOS as an unfavorable (dead, vegetative status, and severe disability) versus favorable (moderate disability and good recovery) outcome. Recent methodological work indicates that analyzing ordinal outcomes as ordinal yields substantial gains over conventional dichotomized outcomes. Therefore it is recommended that future TBI and stroke trials use the original ordinal outcomes and the methods of ordinal analyses.

To further explore the utility of the ordinal GOS in TBI trials, this study extends our previous investigation (Lu et al., 2008) regarding the effect of nondifferential dichotomous GOS misclassification in TBI clinical trials to the impact of GOS misclassification on ordinal analysis in TBI clinical trials. In the previous study, we used a simple sensitivity analysis to explore three patterns of dichotomous GOS misclassification and its impact on the effect size and statistical power. The results suggested that all three simulated patterns of misclassification act to attenuate the treatment effect and reduce the statistical power. In the case of a positive drug effect, misclassification is likely to lead to a conservative estimation of true efficacy.

In this study, we explored the impact of GOS misclassification on the ordinal analysis in TBI trials via probabilistic sensitivity analyses (Fox et al., 2005; Lash and Fink, 2003), using actual TBI patient datasets contained in the

---

[1]Department of Epidemiology and Community Health, and [2]Department of Neurosurgery, Virginia Commonwealth University, Richmond, Virginia.

International Mission for Prognosis And Clinical Trial (IMPACT) database ($n = 9205$; Marmarou et al,. 2007). We examined the impact of GOS misclassification on the primary outcome estimation by analyzing simulated treatment effects in an ordinal analysis of GOS. The conventional 95% confidence intervals and a simulation interval, which accounts for the misclassification and random errors together, were reported. The results will help investigators to better understand the impact of GOS misclassification on the primary outcome estimation in TBI trials in a quantitative manner.

Methods

Patient data

We used patient data from the IMPACT database (Marmarou et al., 2007) as examples of typical selected head-injury populations for late-phase TBI clinical trials. The IMPACT project was an international collaboration linking researchers in Belgium, the Netherlands, the United Kingdom, and the United States of America, which was funded by the National Institutes of Health (NIH), and aimed to develop methodologies to improve the design and analysis of clinical trials of TBI. The IMPACT database contains clinical data on 9205 individual patients with moderate or severe head injury from eight randomized controlled trials (RCTs; $n = 6535$), and three observational epidemiologic studies ($n = 2670$). The patient data from RCTs represents the head-injury population with more restricted inclusion criteria, whereas the data from the observational studies represents the population with more generalized characteristics.

For each individual study in the IMPACT database, 400 subjects were randomly sampled with replacement, as the baseline samples or the placebo group. Another 400 subjects were randomly sampled from each placebo group with replacement, as the treatment group. Thus each dataset was generated representing the general TBI population from RCTs and observational studies.

Simulation of treatment effect

Within each dataset sampled at random, a 10% treatment effect was simulated for the treatment group based on the assumption that the effect of drug treatment followed a proportional odds model (McCullagh, 1980). The common odds ratio (COR) was calibrated so that there would be an overall 10% (COR = 1.5) increase in the proportion of patients with a better outcome in the treatment group.

Simulation of nondifferential GOS misclassification

The GOS includes categories of good recovery (GR), moderate disability (MD), severe disability (SD), vegetative status (VS), and death (D). For this study, the categories of VS and D were combined and analyzed as one category for both logistical and statistical reasons. A few assumptions were made regarding the GOS misclassification. First, in the context of double-blind clinical trials, we assumed that the GOS misclassification was nondifferential, that is, the probability of the misclassification was the same for both treatment and control groups. Second, no misclassification was made for the category of VS, and thus for the combined category of VS and D. Third, the misclassification was made between two adjacent GOS categories only (i.e., between the category of MD and the

two categories [GR and SD] next to it), with the same set of sensitivity and specificity parameters. For example, the misclassification was made between the categories of MD and GR, and between MD and SD, with a 90% sensitivity and specificity. Based on these assumptions, the GOS misclassification was simulated through two independent binary situations, that is, between MD and GR, and between MD and SD, and the overall expected MD was recalculated using the formula as illustrated in Figure 1.

Because the sensitivity and specificity (the classification parameters) are seldom known with certainty, we described a range of misclassification via a trapezoidal distribution (minimum value of 80%, modes of 85 and 95%, and a maximum value of 100%). With the description, we intended to give more emphasis to the error probabilities between 5 and 15% (i.e., between the modes of 85 and 95%), but less to the probabilities between 0 and 5% and between 15 and 20% (i.e., between the minimum value of 80% and the mode of 85%, and between the mode of 95% and maximum value of 100%).

Three patterns of misclassification

Three common patterns of misclassification were simulated to extrapolate the impact of GOS misclassifications on TBI trials, including the random, upward, and downward patterns (Fig. 2). For the random pattern (Fig. 2A), we specified the same trapezoidal distribution for both sensitivity and specificity. With the specification of misclassification, the patients were simulated to be misclassified equally between the two adjacent categories (i.e., between SD and MD, and between MD and GR), for both treatment groups. For the upward pattern (Fig. 2B), we specified the trapezoidal distribution for sensitivity but with perfect specificity. With the specification, the patients were misclassified as experienced better outcomes than the reality for both treatment groups. For example, more patients in the category of SD were misclassified as MD, and more patients in the MD category
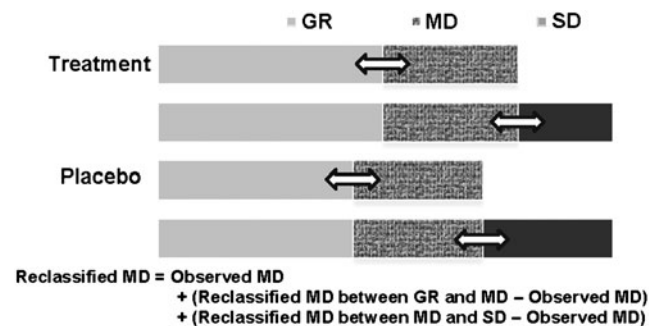

Reclassified MD = Observed MD
+ (Reclassified MD between GR and MD – Observed MD)
+ (Reclassified MD between MD and SD – Observed MD)

FIG. 1. Diagram illustrating the scenario of Glasgow Outcome Scale (GOS) misclassification. To simulate the situation, we assumed that (1) the misclassification can only be made between moderate disability (MD) and two adjacent categories (good recovery [GR] and severe disability [SD]), and (2) no misclassification for the category of vegetative status (VS). Based on these assumptions, the GOS misclassification was simulated through two independent binary situations, that is, between MD and GR, and between MD and SD, and the overall expected MD was recalculated provided the above, which is the observed MD plus the differences between the two independently reclassified MDs and the observed data.
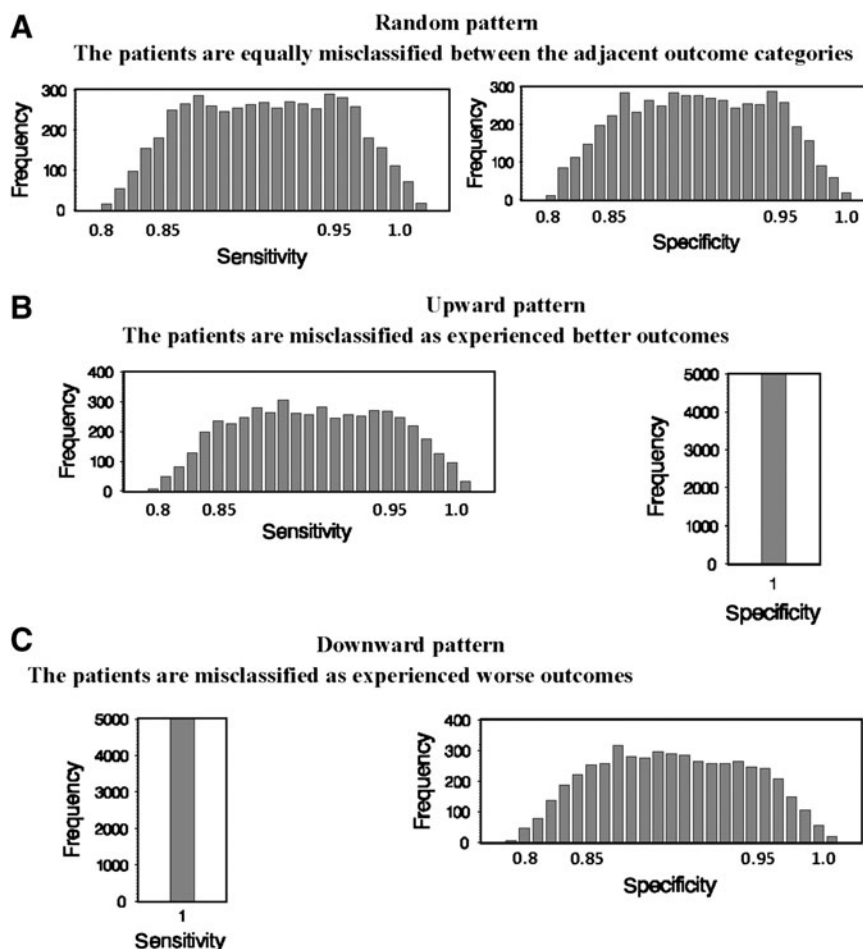
FIG. 2. Graphs showing the probability distributions of the classification parameters. The parameters are used to describe three patterns of misclassification for the study. Panel **A** shows the random pattern for which we specified a trapezoidal distribution (minimum of 80%, modes of 85 and 95%, and a maximum of 100%), for both sensitivity and specificity. Panel **B** shows the upward pattern for which we specified the same trapezoidal distribution for sensitivity, but a perfect specificity. Panel **C** shows the downward pattern, for which we specified the same trapezoidal distribution for specificity, but a perfect sensitivity.

were misclassified as GR, for both treatment groups. On the contrary, for the downward pattern (Fig. 2C), we specified the trapezoidal distribution for specificity, but with perfect sensitivity. Through the specification, the patients were misclassified as experiencing worse outcomes than the reality for both treatment groups. For instance, more patients in the GR category were misclassified as MD, and more patients in the MD category were misclassified as SD for both treatment groups.

Probability sensitivity analysis and output

We used the concept of the probabilistic sensitivity analysis introduced by Lash and Fink (2003), and Fox and colleagues (2005), and modified the SENSMAC (SAS Macros) by Fox and colleagues. The original SENSMAC was generated to provide probabilistic sensitivity analysis to quantify the likely effects of misclassification of a dichotomous outcome, exposure, or covariate. We modified the misclassification section of the SAS macros to explore the impact of GOS misclassification on ordinal analysis in TBI trials. The analysis simulated the data that would have been observed had the misclassified variable been correctly classified, given the sensitivity and specificity

of classification (i.e., the trapezoidal error distribution described in an earlier section). For each study and each pattern of misclassification, the simulation analysis was repeated 5000 times to generate simulation intervals of outcome estimations after correcting for misclassification, as well as correcting for both misclassification and random error. The estimation of GOS at 6 months post-injury was assessed using a proportional odds model and reported via a common odds ratio for the treatment versus the placebo group. No covariate was involved in the analysis. The conventional 95% confidence intervals that accounted for random error only, and simulation intervals that accounted for misclassification and random error, were reported together.

Results

Distribution of GOS at 6 months post-injury

Figure 3 shows the observed 6-month GOS data that were used to explore the impact of GOS misclassification on TBI clinical trials. The outcome datasets were randomly sampled from eight RCTs and three observational studies contained in
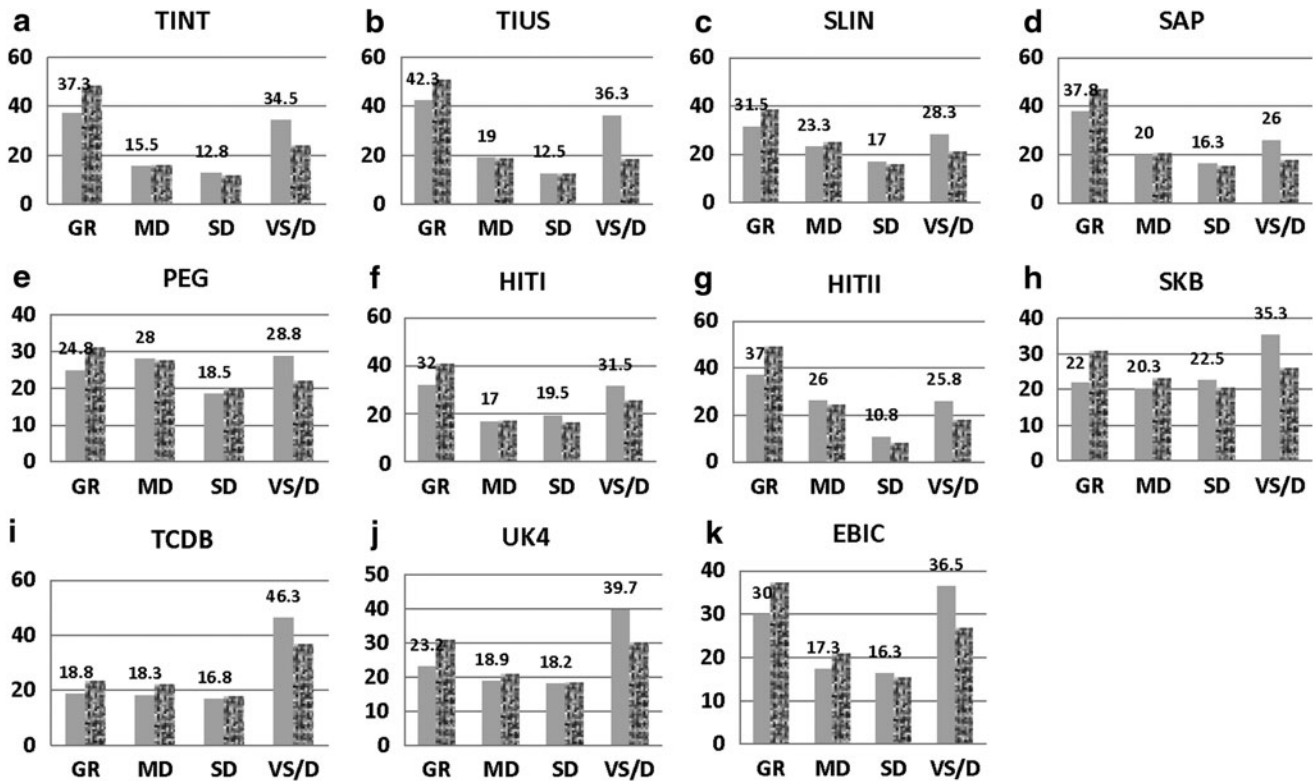
FIG. 3. Graphs showing the observed Glasgow Outcome Scale (GOS) distribution at 6 months post-injury from 11 traumatic brain injury studies contained in the IMPACT database, representing the moderate-to-severe head-injury population. A 10% treatment effect is added for each treatment group via a proportional odds model to symbolize a trial effect. The GOS includes categories of good recovery (GR), moderate disability (MD), severe disability (SD), and combined vegetative status and death (VS/D). The plain pattern represents the placebo group, while the textured pattern represents the treatment group. For each placebo group, the proportions of the individual GOS categories are shown (IMPACT, International Mission for Prognosis And Clinical Trial; TINT, The International Tirilazad Trial; TIUS, The North American Tirilazad Trial; SLIN, The International NMDA Antagonist Selfotel Trial; SAP, The Saphir Study; PEG, The PEGSOD Trial; HITI, The Nimodipine I Trial; HITII, The Nimodipine II Trial; SKB, The Bradycor Trial; TCDB, Traumatic Coma Data Bank; UK4, U.K. Four Center Study; EBIC, The European Brain Injury Consortium Core Data).

the IMPACT database, representing the moderate-to-severe head-injury population with either more restricted inclusion criteria (RCTs), or more generalized characteristics (observational studies). Each data set contains 800 patients with 400 in each arm. A 10% treatment effect was simulated via a proportional odds model for the treatment group to symbolize a trial effect.

In general, most studies had a U-shaped GOS outcome distribution at 6 months post-injury, that is, large proportions of patients had outcomes of either good recovery or combined mortality and vegetative status, and relatively lower percentages of patients had outcomes of moderate or severe disability. The proportions of GR and VS/D among RCTs ranged from 22.0% (The Bradycor Trial, SKB) to 42.3% (The North American Tirilazad Trial, TIUS), and from 25.8% (The Nimodipine II Trial, HITII) to 35.3% (SKB), respectively, while the proportions among the observational studies ranged from 18.8% (Traumatic Coma Data Bank, TCDB) to 30% (The European Brain Injury Consortium Core Data, EBIC), and from 36.5% (EBIC) to 46.3% (TCDB), respectively. The proportions of MD and SD among the RCTs ranged from 15.5% (The International Tirilazad Trial, TINT) to 28% (The PEGSOD Trial, PEG), and from 10.8% (HITII) to 19.5% (The Nimodipine I

Trial, HITI), respectively, whereas the proportion among the observational studies ranged from 17.3% (EBIC) to 18.9% (U.K. Four Center Study, UK4), and from 16.3% (EBIC) to 18.2% (UK4), accordingly.

As expected, the baseline GOS scores at 6 months among the RCTs were better than the outcomes from the observational studies. The proportions of favorable outcomes (GR and MD) were higher among the RCTs, compared with the proportions seen among the observational studies; whereas the mortalities were lower among the RCTs, compared with the mortalities among the observational studies.

Probabilistic sensitivity analysis: Correcting for nondifferential misclassification errors

Table 1 shows the results of the ordinal GOS analyses comparing the estimates of effect ignoring the misclassification, and the probabilistic sensitivity analysis correcting for three patterns of nondifferential misclassification. The estimate of effect ignoring misclassification was performed on each observed dataset, from which a perfect outcome classification was assumed, and the 95% confidence intervals (CI) took account for random errors only. Among all studies, the

TABLE 1. RESULTS OF THE 6-MONTH ORDINAL GLASGOW OUTCOME SCALE[a] ANALYSIS: COMPARISON BETWEEN THE ESTIMATES IGNORING MISCLASSIFICATION AND PROBABILISTIC SENSITIVITY ANALYSIS CORRECTED FOR NONDIFFERENTIAL MISCLASSIFICATION[b]

TBI studies	Randomized controlled trials								Observational studies		
	TINT	TIUS	SLIN	SAP	PEG	HITI	HITII	SKB	TCDB	UK4	EBIC
	Common odds ratios (95% confidence intervals accounting for random error only)										
Estimates ignoring misclassification	1.62 (1.25,2.09)	1.46 (1.13,1.89)	1.41 (1.09,1.81)	1.51 (1.17,1.95)	1.35 (1.06,1.71)	1.42 (1.11,1.82)	1.61 (1.23,2.08)	1.57 (1.22,2.00)	1.42 (1.12,1.85)	1.49 (1.16,1.92)	1.50 (1.18,1.91)
Estimates corrected for misclassification	Median common odds ratios (2.5 and 97.5 percentiles of the simulation intervals accounting for misclassification and random error)										
Random misclassification[c] (i.e., the patients are misclassified equally between the adjacent categories)	1.67 (1.29,2.18)	1.51 (1.15,1.98)	1.45 (1.11,1.87)	1.55 (1.18,2.04)	1.39 (1.06,1.82)	1.43 (1.11,1.86)	1.71 (1.31,2.26)	1.60 (1.23,2.07)	1.43 (1.10,1.86)	1.51 (1.15,1.96)	1.50 (1.17,1.93)
Upward misclassification[d] (i.e., the patients are misclassified as experienced better outcomes)	1.66 (1.28,2.17)	1.50 (1.14,1.96)	1.43 (1.11,1.85)	1.53 (1.18,1.99)	1.36 (1.06,1.76)	1.44 (1.12,1.85)	1.67 (1.29,2.19)	1.59 (1.23,2.06)	1.43 (1.11,1.87)	1.50 (1.17,1.95)	1.50 (1.16,1.66)
Downward misclassification[e] (i.e., the patients are misclassified as experienced worse outcomes)	1.43 (1.11,1.84)	1.28 (0.97,1.68)	1.12 (0.87,1.45)	1.24 (0.96,1.61)	1.02 (0.78,1.31)	1.13 (0.88,1.45)	1.46 (1.11,1.89)	1.13 (0.87,1.48)	1.12 (0.87,1.45)	1.16 (0.90,1.51)	1.21 (0.95,1.57)

[a]The ordinal Glasgow Outcome Scales (GOS) are categorized as: combined dead and vegetative status, severe disability, moderate disability, and good recovery.
[b]Nondifferential misclassification is assumed between the two adjacent outcome categories, only except for the combined category of dead and vegetative status, for both treated and placebo groups.
[c]Sensitivity and specificity are drawn from a trapezoidal distribution, with a minimum of 80% modes of 85% and 95%, and a maximum of 100% for each.
[d]Sensitivity is drawn from a trapezoidal distribution, with a minimum of 80% modes of 85% and 95%, and a maximum of 100%, but specificity is defined as 100%.
[e]Sensitivity is defined as 100%, but specificity is drawn from a trapezoidal distribution, with a minimum of 80% modes of 85% and 95%, and a maximum of 100%.
Table 1 shows the results of the ordinal GOS analyses on eight traumatic brain injury clinical trials and three observational studies, comparing the estimates of effect ignoring the misclassification, and the probabilistic sensitivity analysis correcting for three patterns of nondifferential misclassification. The proportional odds model is applied to estimate the common odds ratios of more favorable outcomes, as compared between the treatment and placebo groups.
TINT, The International Tirilazad Trial; TIUS, The North American Tirilazad Trial; SLIN, The International NMDA Antagonist Selfotel Trial; SAP, The Saphir Study; PEG, The PEGSOD Trial; HITI, The Nimodipine I Trial; HITII, The Nimodipine II Trial; SKB, The Bradycor Trial; TCDB, Traumatic Coma Data Bank; UK4, U.K. Four Center Study; EBIC, The European Brain Injury Consortium Core Data.

723

common odds ratios of more favorable outcome, as compared between the treatment and placebo groups, ranged from 1.35 (PEG, 95% CI 1.06,1.71) to 1.62 (TINT, 95% CI 1.25,2.09).

Misclassification with random and upward patterns

The probabilistic sensitivity analysis, correcting for the misclassification with a random pattern (i.e., patients were misclassified equally between the two adjacent categories for both treatment groups), was demonstrated by a trapezoidal error distribution specified for sensitivity and specificity, with a minimum of 80%, modes of 85% and 95%, and a maximum of 100% for each. The 95% simulation intervals, which account for misclassification and the corresponding median estimates, moved upward slightly compared to the results of the conventional approach (i.e., ignoring misclassification and accounting for random error only) for all studies. Consequently, the overall 95% simulation intervals, that account for both misclassification and random errors moved upward, ranging from 1.06 to 1.82 (PEG) to 1.29 to 2.18 (TINT), and the corresponding median estimates ranged from 1.39 to 1.67.

Given the specified sensitivity (minimum of 80%, modes of 85 and 95%, and a maximum of 100%), and specificity (100%) parameters, the analysis results, correcting for the misclassification with an upward pattern (i.e., patients were misclassified as experienced better outcomes for both treatment groups), were similar to the results seen with a random pattern. The 95% simulation intervals and the corresponding median estimates (correcting for misclassification only) moved upward slightly compared to the results of the conventional approach (i.e., ignoring misclassification and accounting for random error only) for all studies. Accordingly, the overall 95% simulation intervals that account for both misclassification and random error moved upward, ranging from 1.06 to 1.76 (PEG) to 1.28 to 2.17 (TINT), and the corresponding median estimate ranged from 1.36 to 1.66. Thus, if a random or upward pattern of nondifferential misclassification existed within the specified classification error ranges, the GOS at 6 months post-injury from the observed datasets would have been underestimated by a small degree.

Misclassification with a downward pattern

In contrast, given the specification of the sensitivity (100%) and specificity (minimum of 80%, modes of 85 and 95%, and a maximum of 100%) parameters, the analysis results, correcting for the misclassification with a downward pattern (i.e., patients were misclassified as experienced worse outcomes for both treatment groups), were different from the results correcting for the random and upward misclassification. The 95% simulation limits and the corresponding median estimates (correcting for misclassification only) moved downward compared to the results of the conventional approach (i.e., ignoring misclassification and accounting for random error only) for all studies. So did the overall 95% simulation limits that account for misclassification and random errors for all studies. For the downward pattern, the overall 95% simulation intervals ranged from 0.78 to 1.31 (PEG) to 1.11 to 1.84 (TINT), and the corresponding median estimates ranged from 1.02 to 1.43. Therefore, if a downward pattern of nondifferential misclassification existed within the assumed error ranges, the ordinal outcome from the observed datasets would have been inflated.

Discussion

We explored the impact of nondifferential GOS misclassification on ordinal analysis in TBI clinical trials via a probabilistic sensitivity analysis using TBI patient datasets from the IMPACT database. The analysis involved reconstructing the data that would have been observed had the misclassified variable been correctly classified, given the sensitivity and specificity of classification. We examined the impact of GOS misclassification on the primary outcome estimation by analyzing simulated treatment effects in an ordinal analysis of GOS. We have demonstrated that nondifferential misclassification could produce uncertainties on the ordinal GOS analysis in TBI trials. For instance, our simulation results showed that (1) given a specification of a minimum of 80%, modes of 85% and 95%, and a maximum of 100%, for both sensitivity and specificity (i.e., a random pattern of misclassification for which patients were misclassified equally between the two adjacent outcome categories for both treatment groups), and (2) given the same trapezoidal-distributed sensitivity but a perfect specificity (i.e., an upward pattern of misclassification for which patients were misclassified as experienced better outcomes for both treatment groups), the misclassification would have caused an ordinal GOS slightly underestimated in the observed datasets. In another scenario, given the same trapezoidal-distributed specificity but a perfect sensitivity (i.e., a downward pattern of misclassification for which patients were misclassified as experienced worse outcomes for both treatment groups), the misclassification would have resulted in an inflated GOS estimation.

Outcome misclassification in TBI clinical trials

In practice, it is highly possible that the primary outcomes such as GOS and GOS Extended (GOSE) could have been misclassified to some extent in the TBI trials. Various researchers have investigated misclassification and inter-observer variation of the TBI outcome measures, and in general found that the variation does exist (Anderson et al., 1993; Brooks et al., 1986; Maas et al., 1983; Marmarou, 2001; Pettigrew et al., 1998; Scheibel et al., 1998; Teasdale et al., 1998; Wilson et al., 1998,2002,2007). The reported overall disagreement in GOS assessments ranged from 8% (Wilson et al., 1998) to 30% (Brooks et al., 1986), whereas the disagreement in GOSE ratings ranged from 22% (Wilson et al., 1998) to 41% (Wilson et al., 2007). When the overall disagreement in GOS assessment (collapsed from GOSE) from the study was broken down into individual categories (Wilson et al., 2007), the disagreement levels in rating the categories of SD, MD, and GR, between an expert and the untrained investigators, were 29.5%, 53.3%, and 35%, respectively.

Three patterns of nondifferential misclassification

Our sensitivity analyses are based on the GOS misclassification scenarios that are common to clinical practice. For example, Marmarou (2001) conducted a study among 34 American Brain Injury Consortium members to ascertain the reliability of the GOS rating. The results showed that the rating for 20.6% of moderately-disabled patients was shifted to the good recovery category, and 32.3% of severely-disabled patients were rated as moderately disabled. An upward shift of outcome assignment had been previously reported (Anderson et al., 1993), and is a likely result of the optimism of the

patient's primary care providers, who compared the improved outcome to the serious condition seen immediately after injury, rather than to the healthy pre-injury status. Conversely, a rigid application of the criteria from the structured interview or questionnaires by research workers tends to allocate patients to lower outcome categories (Teasdale et al., 1998; Wilson et al., 1998). Therefore, nondifferential misclassification may be found in either the upward or downward direction, based on different clinical scenarios.

Correlation between the nondifferential misclassification and the probabilities of GOS categories

In this sensitivity analysis, it appears that the impact of nondifferential GOS misclassification on ordinal analysis in TBI trials is less significant, compared with the effect of the dichotomous GOS misclassification situation reported previously (Lu et al., 2008). This is likely due to the probabilities or the prevalence of GOS categories that are misclassified. The correlations between the nondifferential misclassification and three examples of GOS category probability sets are given in Table 2. We propose that three GOS category probability sets (i.e., equal probability, the U-shaped distribution, and single dominant category) reflect the true outcome distribution, whereas the GOS assessment is done with errors. The classification errors are illustrated via a simple model, in which 20% of patients in category GR are classified as being in MD, 20% MD being in GR, 20% of patients in category MD being in SD, and 20% SD being in MD, for both placebo and treatment groups. As a result, the true category probabilities given at the beginning of each case (rows of "True outcome") are transformed by misclassification into the observed probabilities (rows of "Random misclassification").

The results from our examples confirmed that the effect of misclassification on the cases of equal probability and U-shaped GOS distribution is relatively small. However, given the same moderate classification error rate (e.g., 20%), and

overall treatment effect (e.g., 10%), the random misclassification caused the true outcome to be substantially underestimated in a single dominant ordinal GOS scenario, and the true outcome difference between the placebo and treatment groups was reduced from 10% (common odds ratio = 1.5) to 7.4% (common odds ratio = 1.35) in our example. The scenario is similar with the effect of misclassification on the binary GOS data that poses a relatively large difference in the probabilities between the favorable and unfavorable outcome categories. Thus, the impact of misclassification will likely be less sensible in the equal probability and the U-shaped GOS distributions, as observed in the 11 TBI studies presented in Table 1. The results of this demonstration are consistent with the examples given by Whitehead (1993).

Advantages and limitations of the probabilistic sensitivity analysis

Taken together, the scenario and the simulated error intervals extrapolated by this study are in accordance with previous study results. Based on the ranges of inter-observer variation observed from previous TBI studies, we used a trapezoidal distribution to describe the values of the misclassification parameters (i.e., the values of sensitivity and specificity). Overall, the error distribution is specified by four points: the lower (80%) and upper bounds (100%), and the lower (85%) and upper (95%) modes. With the description, the rate of potential misclassification between the two adjacent GOS categories ranges from a minimum of 0% (sensitivity or specificity equal 100%), to a maximum of 20% (sensitivity or specificity equal 80%). Within the range, a more plausible range of misclassification is between 5 and 15% (sensitivity or specificity ranges from 85% to 95%). Thus, unlike the simple sensitivity analyses, the results from this probabilistic sensitivity analysis provide a sense of central tendency of the corrected ordinal GOS estimate. The results also provide a measure of uncertainty in the corrected estimate, as portrayed by the simulation intervals that include both misclassification

TABLE 2. THE CORRELATIONS BETWEEN THE NONDIFFERENTIAL MISCLASSIFICATION AND THREE PROBABILITY SETS OF THE GLASGOW OUTCOME SCALE

| Cases | Analysis | Placebo (n = 400) | | | | Treatment (n = 400) | | | | Common odds ratios[c] |
		GR	MD	SD	D/VS	GR	MD	SD	D/VS	
Equal probability	True outcome[a]	0.25	0.25	0.25	0.25	0.33	0.27	0.22	0.18	1.50
	20% Random misclassification[b]	0.25	0.25	0.25	0.25	0.32	0.27	0.23	0.18	1.44
U-shaped distribution	True outcome	0.35	0.15	0.15	0.35	0.45	0.15	0.14	0.27	1.50
	20% Random misclassification	0.31	0.19	0.15	0.35	0.39	0.21	0.14	0.27	1.46
Single dominant category	True outcome	0.20	0.50	0.20	0.10	0.27	0.51	0.15	0.07	1.50
	20% Random misclassification	0.26	0.38	0.26	0.10	0.32	0.39	0.22	0.07	1.35

[a]The probabilities of the individual Glasgow Outcome Scale (GOS) categories were assumed as true, with a 10% treatment effect added to the treatment group based on the proportional odds model assumption.

[b]For the random misclassification, a 20% rate exchange between the categories of SD (severe disability) and MD (moderate disability), and between the categories of MD and GR (good recovery), was applied for both treatment groups. No misclassification was made for the category of D/VS (combined categories of dead and vegetative status).

[c]The common odds ratio was estimated via a proportional odds model.

Table 2 shows the correlations between the nondifferential misclassification and three examples of GOS probability sets. In the illustration, we propose that these probability sets are the true outcome distribution; however, the GOS assessment is done with errors. The classification errors are illustrated via a simple model, in which 20% of patients in the category of good recovery (GR) are classified as being in moderate disability (MD), 20% MD being in GR, 20% of patients in category MD being in severe disability (SD), and 20% of SD being in MD, for both the placebo and treatment groups. As a result, the true category probabilities given at the beginning of each case (the rows of "True outcome") are transformed by misclassification into the observed probabilities (the rows of "Random misclassification").

and random errors. More significantly, our simulation study was based on data from eight major Phase III trials in TBI and three observational TBI studies. As such, we believe that our findings may be applicable to a wide range of trials in TBI.

It should be pointed out that similarly to all simulation studies, the main limitation of this study was that the distribution of the assumed misclassification parameter may be arbitrary, which could lead to different distributions of the corrected analysis. Furthermore, the informed sensitivity analysis may be limited by the absence of any sense of weight to yield various results, such as the rate of misclassification between GR and MD or between MD and SD. In practice, the rate of misclassification may well be different between GR (good recovery) and MD (moderate disability) versus between MD and SD (severe disability).

In conclusion, the probabilistic sensitivity analysis from this study suggests that given the moderate classification error ranges as specified by this study, the impact of nondifferential GOS misclassification on ordinal analysis in TBI clinical trials is likely to be small, compared with the effect of binary GOS misclassifications in TBI trials. The findings were consistent across eight major Phase III TBI trials and three observational studies. The results underline the fact that the ordinal GOS analysis may gain from both the statistical efficiency, as suggested by several recent TBI and stroke studies, and a relatively smaller impact of the misclassification, compared with the conventional binary GOS analysis. Nevertheless, outcome assessment following TBI is a complex problem. The assessment quality may be influenced by many factors. All possible aspects must be considered to ensure the consistency and reliability of the assessment, and to optimize the success of the trial.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

Anderson, S.I., Housley, A.M., Jones, P.A., Slattery, J., and Miller, J.D. (1993). Glasgow Outcome Scale: an inter-rater reliability study. Brain Inj. 7, 309–317.

Bath, P.M., Geeganage, C., Gray, L.J., Collier, T., and Pocock, S. (2008). Use of ordinal outcomes in vascular prevention trials: comparison with binary outcomes in published trials. Stroke 39, 2817–2823.

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.

Fox, M.P., Lash, T.L., and Greenland, S. (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int. J. Epidemiol. 34, 1370–1376.

Jennett, B., and Bond, M. (1975). Assessment of outcome after severe brain damage. Lancet 1, 480–484.

Lash, T.L., and Fink, A.K. (2003). Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology (Cambridge, Mass.) 14, 451–458.

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.

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.

Marmarou, A. (2001). *Head Trauma: Basic, Preclinical, Clinical Direction*, 1st ed. Wiley: New York, p. 15.

Marmarou, A., Lu, J., Butcher, I., McHugh, G.S., Mushkudiani, N.A., Murray, G.D., Steyerberg, E.W., and Maas, A.I. (2007). IMPACT database of traumatic brain injury: design and description. J. Neurotrauma 24, 239–250.

McCullaph, P. (1980). Regression-models for ordinal data. J. R. Statist. Soc. Ser. B Methodological 42, 109–142.

McHugh, G.S., Butcher, I., Steyerberg, E.W., Marmarou, A., Lu, J., Lingsma, H.F., Weir, J., Maas, A.I., and Murray, G.D. (2010). A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. Clinical Trials (London, England) 7, 44–57.

Optimising Analysis of Stroke Trials (OAST) Collaboration, Bath, P.M., Gray, L.J., Collier, T., Pocock, S., and Carpenter, J. (2007). Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke 38, 1911–1915.

Pettigrew, L.E., Wilson, J.T., and Teasdale, G.M. (1998). Assessing disability after head injury: improved use of the Glasgow Outcome Scale. J. Neurosurg. 89, 939–943.

Scheibel, R.S., Levin, H.S., and Clifton, G.L. (1998). Completion rates and feasibility of outcome measures: experience in a multicenter clinical trial of systemic hypothermia for severe head injury. J. Neurotrauma 15, 685–692.

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.

Whitehead, J. (1993). Sample size calculation for ordered categorical data, Stat. Med. 12, 2257–2271.

Wilson, J.T., Edwards, P., Fiddes, H., Stewart, E., and Teasdale, G.M. (2002). Reliability of postal questionnaires for the Glasgow Outcome Scale. J. Neurotrauma 19, 999–1005.

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.

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.

Address correspondence to:
Juan Lu, M.D., Ph.D.
Virginia Commonwealth University
Department of Epidemiology and Community Health
Department of Neurosurgery
P.O. Box 980212
Richmond, VA 23298-0212

Email: jlu1@vcu.edu