Received: 20 July 2022 | Revised: 14 June 2023 | Accepted: 1 August 2023

DOI: 10.1002/sim.9877

RESEARCH ARTICLE

Statistics in Medicine WILEY

Selection of a statistical analysis method for the Glasgow Outcome Scale-Extended endpoint for estimating the probability of favorable outcome in future severe TBI clinical trials

Yu Wang[1,2] | **Sharon D. Yeatts**[3] | **Renee' H. Martin**[3] | **Robert Silbergleit**[4] | **Gaylan L. Rockswold**[5] | **William G. Barsan**[4] | **Frederick K. Korley**[4] | **Sarah Rockswold**[5] | **Byron J. Gajewski**[1]

[1]Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas City, Kansas, USA

[2]Global Biometrics & Data Sciences, Bristol Myers Squibb, Lawrenceville, New Jersey, USA

[3]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

[4]Department of Emergency Medicine, University of Michigan, Ann Arbor, Michigan, USA

[5]Department of Neurosurgery, University of Minnesota, Hennepin County Medical Center, Minneapolis, Minnesota, USA

Correspondence
Yu Wang, Department of Biostatistics & Data Science, University of Kansas Medical Center, Robinson 5028, 3901 Rainbow Blvd., Kansas City, KS 66160, USA and Global Biometrics & Data Sciences, Bristol Myers Squibb, Lawrenceville, NJ 08540, USA.
Email: cindywanghuster@gmail.com

Funding information
National Institutes of Health

The Glasgow outcome scale-extended (GOS-E), an ordinal scale measure, is often selected as the endpoint for clinical trials of traumatic brain injury (TBI). Traditionally, GOS-E is analyzed as a fixed dichotomy with favorable outcome defined as GOS-E ≥ 5 and unfavorable outcome as GOS-E < 5. More recent studies have defined favorable vs unfavorable outcome utilizing a sliding dichotomy of the GOS-E that defines a favorable outcome as better than a subject's predicted prognosis at baseline. Both dichotomous approaches result in loss of statistical and clinical information. To improve on power, Yeatts et al proposed a sliding scoring of the GOS-E as the distance from the cutoff for favorable/unfavorable outcomes, and therefore used more information found in the original GOS-E to estimate the probability of favorable outcome. We used data from a published TBI trial to explore the ramifications to trial operating characteristics by analyzing the sliding scoring of the GOS-E as either dichotomous, continuous, or ordinal. We illustrated a connection between the ordinal data and time-to-event (TTE) data to allow use of Bayesian software that utilizes TTE-based modeling. The simulation results showed that the continuous method with continuity correction offers higher power and lower mean squared error for estimating the probability of favorable outcome compared to the dichotomous method, and similar power but higher precision compared to the ordinal method. Therefore, we recommended that future severe TBI clinical trials consider analyzing the sliding scoring of the GOS-E endpoint as continuous with continuity correction.

KEYWORDS

Bayesian design, clinical trials, Glasgow outcome scale-extended, sliding scoring, traumatic brain injury

1 | INTRODUCTION

Traumatic brain injury (TBI) is a leading cause of death and disability worldwide.[1] Despite decades of clinical trials, there is no specific pharmacologic agent available to improve outcomes after TBI. Appropriate statistical methods are critical

 © 2023 John Wiley & Sons Ltd.

components of clinical trials, and more efficient statistical design and analysis are essential to improving the likelihood of success of clinical trials in TBI.[2]

The Glasgow outcome scale-extended (GOS-E) is an ordinal scale most often selected as the primary endpoint in traumatic brain injury clinical trials. The GOS-E rates patient functional status into one of eight categories. A GOS-E score of 1 indicates death, 2 indicates a vegetative state, 3 and 4 indicates lower and upper severe disability, respectively, 5 and 6 indicates lower and upper moderate disability respectively, and 7 and 8 indicates lower and upper good recovery, respectively. A straightforward analysis of the GOS-E is based on the fixed dichotomy, which defines a favorable outcome as GOS-E ≥ 5 and unfavorable outcome as GOS-E ≤ 5.[3] The fixed dichotomy is often criticized for its statistical inefficiency since it usually requires a larger sample size to demonstrate efficacy than other approaches.[4] In addition, the analysis is clinically problematic, as a score of 5 (lower moderate disability) is considered identical to a score of 8 (upper good recovery), and a 4 (upper severe disability) is considered the same as a score of 1 (death). Furthermore, the fixed dichotomy approach does not account for the initial severity of the injury, or the predicted prognosis.

Accounting for baseline prognosis is a common approach to improve power in TBI clinical trials.[5] Proposed by Murray et al,[6] the sliding dichotomy of the GOS-E defines a favorable outcome as a better outcome than predicted based on patient's injury characteristics. The rationale is that any improvement from an expected poor outcome relative to the baseline severity is clinically meaningful, even if it does not meet the threshold for favorable outcome based on a fixed cut-off point ($\geq$5). However, the limitation of the sliding dichotomy approach is the loss of statistical information and reduced power due to the collapse of adjacent categories,[7,8] and further improvements are needed to advance analysis methods for TBI clinical trials.

As an alternative to binary analysis, analysis that maintains the ordinal nature of the data, such as a proportional odds (PO) model, has been shown to be more statistically efficient than the dichotomization approaches.[9-11] Dijkland et al showed that ordinal analysis based on a proportional odds model offers higher power than dichotomous analysis based on logistic regression or continuous analysis based on linear regression for stroke trial designs with the utility-weighted modified Rankin scale (UW-mRS) as the primary outcome. McHugh et al demonstrated that the proportional odds model allows sample sizes to be reduced without loss of statistical power compared to either the fixed or sliding dichotomy,[11] but suggested that its efficiency might be at odds with its clinical appeal. To exploit the ordinal nature of the GOS-E, a recent paper by Yeatts et al proposed a sliding scoring as the numeric distance (from -5 to 4) from the cutoff for favorable outcomes. This work treated the sliding score as continuous based on a normal distribution and estimated the probability of favorable outcome as the probability equal to or greater than zero. This approach demonstrated an improvement in statistical power but did not account for the potential bias associated with treatment of the ordinal score as continuous based on the normal approximation without continuity correction.[12]

In this investigation, we proposed analyzing the sliding scoring of the GOS-E as truly ordinal. We used data from a large phase III TBI randomized control trial (The Progesterone for the Treatment of Traumatic Brain Injury Experimental Clinical Treatment Trial, ProTECT-III) for the prospective Bayesian design using an ordinal model and evaluated the trial operating characteristics including type I error rate and power when compared to a design using dichotomous outcomes or a continuous model with continuity correction, and compare the efficiency of estimates across the candidate designs in term of bias, variance and mean squared error (MSE). As an illustrative example, we applied dichotomous, continuous, and ordinal models to the ProTECT-III data and compared the posterior mean with 95% credible intervals.

Another novelty of this research is to bridge the analysis between ordinal data and time-to-event (TTE) data. We presented a one-to-one correspondence between the proportional odds model applied on ordinal data and a piecewise exponential model used for TTE data under the proportionality assumptions. The TTE outcome has never been repurposed to approximate an ordinal outcome, though they both hold the proportionality assumption either for the odds ratio or hazard ratio. This connection allows us to use existing Bayesian software for TTE analyzes and compare to a conventional tool.

The article is organized as follows: Sections 2.1 and 2.2 describe the motivation and cover the three statistical models (dichotomous, continuous, and ordinal). Section 2.3 describes the trial simulations including data generating process and hypothetical scenarios. Section 2.4.1 introduces the design operating characteristics to be investigated, as well as details on model calibration to control for the type I error rate. Section 2.4.2 describes the measures used to evaluate the accuracy and precision of Bayesian estimation. The simulation results are summarized in Section 3, and an application to real trial data is provided in Section 4. In Section 5 and 6, we note conclusions from our analysis, discuss its innovations and limitations, and propose the future directions of our research.

2 METHODS

2.1 Motivation

2.1.1 Motivating trials

The motivating trials are the hyperbaric oxygen brain injury treatment trial (HOBIT) and brain oxygen optimization in severe TBI phase-3 (BOOST-3).

HOBIT (NCT02407028) is a phase II Bayesian adaptive clinical trial designed to select the optimal dose regimen of hyperbaric oxygen treatment. Subjects are randomized via covariate adaptive and response adaptive randomization to one of seven therapy arms (pressurized hyperbaric oxygen over 5 days and/or normobaric hyperoxia in four of the arms) or a control arm (standard of care) to treat severe TBI. The primary model is a Bayesian hierarchical EMAX dose-response model.[13]

BOOST-3 (NCT03754114) is a phase III randomized clinical trial to determine the comparative effectiveness of two treatment arms, a strategy guided by treatment goals based on both intracranial pressure (ICP) and brain tissue oxygen ($PbtO_2$) as compared to a strategy guided by treatment goals based on ICP monitoring alone, for monitoring and treating patients with severe TBI in the intensive care unit.[14] Subjects are randomized via equal allocation ratio using a covariate adaptive randomization scheme adjusting for site and severity, as defined by the IMPACT Core model probability of poor outcome.[15] The IMPACT core model is a logistic model derived based on age, Glasgow coma scale (GCS) score, pupil reactivity using an external dataset for validation. Predicted probabilities of poor outcome are calculated from IMPACT core model.[15]

2.1.2 Selection of analysis method

The primary analysis method for both HOBIT and BOOST-3 trials use a sliding dichotomy of the GOS-E (eg, as shown for HOBIT in Table 1), and the primary statistical analysis models were based on binomial distribution. As shown in Table 1, subjects could be either rated as 1 (favorable recovery) or 0 (unfavorable recovery). Although the sliding dichotomy tailors the definition of favorable outcome according to injury severity, all scores meeting the criteria of favorable outcome are considered the same. As an extreme example of this limitation, for a subject having a high probability of poor outcome (0.56-1), an outcome of upper severe disability is considered equally favorable as an outcome of upper good recovery. From a clinical and statistical perspective, this results in a loss of information, and patients and clinicians would assign higher utility to the near perfect outcome.

Yeatts et al proposed the sliding scoring (Table 2) of the GOS-E as the distance from the cutoff for favorable/unfavorable outcome. As shown in Table 2, subjects could be rated from −5 to 4 based on an intersection of the GOS-E raw score and the baseline prognosis. Yeatts et al analyzed the ordinal endpoint as if it were continuous, and subsequently calculated the probability of favorable outcome as $p_j = 1 - \Phi\left(-\frac{\mu_j}{\sigma}\right)$ without continuation correction, which could lead to potential bias.

We propose analyzing the sliding scoring of the GOS-E as truly ordinal and compare the ordinal analysis method to the dichotomous and continuous analysis methods under the Bayesian framework. The aim of this article is to evaluate the

TABLE 1 Dichotomy of the Glasgow outcome scale-extended, as used in HOBIT

Predicted probability of poor outcome, as derived from the IMPACT core model	Glasgow outcome scale-extended						
	Upper good recovery (8)	Lower good recovery (7)	Upper moderate disability (6)	Lower moderate disability (5)	Upper severe disability (4)	Lower severe disability (3)	Vegetative state or death (1 or 2)
0-0.20	1	1	0	0	0	0	0
0.21-0.40	1	1	1	0	0	0	0
0.41-0.55	1	1	1	1	0	0	0
0.56-1	1	1	1	1	1	0	0

TABLE 2 Ordinal sliding scoring of the Glasgow outcome scale-extended.

Predicted probability of poor outcome, as derived from the IMPACT core model	Glasgow outcome scale-extended						
	Upper good recovery (8)	Lower good recovery (7)	Upper moderate disability (6)	Lower moderate disability (5)	Upper severe disability (4)	Lower severe disability (3)	Vegetative state or death (1 or 2)
0-0.20	1	0	−1	−2	−3	−4	−5
0.21-0.40	2	1	0	−1	−2	−3	−4
0.41-0.55	3	2	1	0	−1	−2	−3
0.56-1	4	3	2	1	0	−1	−2

different statistical analysis methods for the GOS-E based on the estimates for the probability that an individual subject has a favorable outcome and make recommendations for the prospective design of TBI clinical trials. We first investigate the operating characteristics (eg, type I error, power) of prospective Bayesian designs via simulation. Then we evaluate the performance (eg, bias, variance, MSE) of the Bayesian posterior estimates of the probability of favorable outcome obtained.

2.2 | Statistical models

Let Y be the 10-category ordinal outcome defined as the distance (−5 to 4) from the cutoff point based on the sliding scoring of the GOS-E. The probability that an individual subject has a favorable outcome, $p_d = P(Y_d \geq 0), d = 1, 2$ is modeled for each arm where Y_d is the ordinal outcome for each of the two arms. Different statistical models will be considered to analyze the same indicator $p_d, d = 1, 2$ through different model parameterizations and distributions of the outcome variable.

2.2.1 | Dichotomous model

We considered a Bayesian logistic model that dichotomizes the ordinal outcome Y_d as

$$Z_d = \begin{cases} 1, & \text{if } Y_d \geq 0, \\ 0, & \text{if } Y_d < 0, \end{cases}$$

where

$Z_d \sim \text{Bernoulli}(p_d), d = 1, 2$:

$$\text{logit}(p_d) = \log\left(\frac{p_d}{1 - p_d}\right) = \theta_d,$$

where p_d is the probability of favorable outcome for control arm ($d = 1$) or treatment arm ($d = 2$), and θ_d is the corresponding log-odds.

We imposed non-informative priors (approximately uniform) on the log-odds θ_d for each arm:

$$\theta_d \sim N\left(\mu_d, v_d^2\right), d = 1, 2,$$

where $\mu_d = 0$. We investigated the impact of different choices of SD $v_d = 0.5, 1, 1.82, 2, 5, 10$ and found that the posterior estimates of the log-odds θ_d were similar when $v_d \geq 1.82$. As we have no prior information on the log-odds θ_d, we decided on a non-informative prior of $N\left(0, 1.82^2\right)$.

2.2.2 | Continuous model

We considered a Bayesian linear model for continuous analysis. The ordinal endpoint Y_d is analyzed as continuous based on normal approximation:

$$Y_d \sim N\left(\eta_d, \sigma^2\right), d = 1, 2,$$

$$\eta_d \sim N\left(\mu_d, \tau_d^2\right), d = 1, 2.$$

We impose a weakly informative prior on η_d:

$$\eta_d \sim N\left(0, 10^2\right), d = 1, 2,$$

$$\sigma^2 \sim \text{Inverse} - \text{Gamma}(\alpha, \beta) = \text{Scaled} - \text{inverse} - X^2\left(\frac{\sigma_n}{2}, \frac{\sigma_\mu^2 \sigma_n}{2}\right),$$

where weight $\sigma_n = 2$ and scale $\sigma_\mu = 1$. As the endpoint Y_d has a range of −5 to 4, the above prior η_d is weakly-informative. For the prior imposed on σ^2, we investigated the impact of different combinations of weight σ_n and scale σ_μ on the type I error rate, and weight $\sigma_n = 2$ and scale $\sigma_\mu = 1$ were chosen to control for the type I error rate at the nominal level of 2.5%.

To account for the fact that we use a continuous (normal) probability distribution to approximate a discrete (binomial) probability distribution, we applied a continuity correction to obtain the probability of a favorable outcome for each arm:

$$p_{\text{d}} = \Pr\left(Y \geq -0.5 | \eta_d, \sigma^2\right) = \int_{-0.5}^{\infty} \phi\left(y | \eta_d, \sigma^2\right) dy, d = 1, 2.$$

2.2.3 | Ordinal model

Proportional odds model

Let $W = Y + 6$ denote the 10-category ordinal outcome ranging from 1 to 10. We considered the PO model for the ordinal endpoint W with J categories as ordinal: $\pi_j = P(W = j|\boldsymbol{x})$ with

$$\sum_{j=1}^{J} \pi_j = 1$$

- The cumulative probabilities: $P(W \leq j|\boldsymbol{x}) = \pi_1(\boldsymbol{x}) + \ \ldots \ + \pi_j(\boldsymbol{x}), j = 1, \ \ldots \ J$
- The exceedance probabilities $P(W > j|\boldsymbol{x}) = \pi_{j+1}(\boldsymbol{x}) + \ \ldots \ + \pi_J(\boldsymbol{x}), j = 1, \ \ldots \ J$

The proportional odds model is a set of cumulative logits in a single parsimonious model:

$$\text{logit}\left\{P(W \leq j|\boldsymbol{x})\right\} = \log\left\{\frac{P(W \leq j|\boldsymbol{x})}{1 - P(W \leq j|\boldsymbol{x})}\right\} = \alpha_j + \boldsymbol{\beta}^T \boldsymbol{x}, j = 1, \ \ldots \ J - 1.$$

Or equivalently,

$$\text{logit}\left\{P(W > j|\boldsymbol{x})\right\} = \log\left\{\frac{P(W > j|\boldsymbol{x})}{1 - P(W > j|\boldsymbol{x})}\right\} = -\alpha_j - \boldsymbol{\beta}^T \boldsymbol{x}, j = 1, \ \ldots \ J - 1,$$

where we imposed a weakly informative prior on the regression coefficients:

$$\alpha_j, \beta \sim N\left(0, 10^2\right), j = 1, \ \ldots \ J - 1.$$

α_j represents the cumulative log odds of being in category j or lower for those in the control group (set as the reference) as well as the threshold of the latent continuous data generating process. β is the cumulative log-odds ratio for the category j comparing the treatment to the control group. β also represents the shift of the threshold on the latent continuous process for the treatment relative to the control. The proportionality assumption implies that the shift of the threshold for each of the categories is identical. Note that in the model parameterization based on the cumulative probabilities $P(W \leq j|\boldsymbol{x})$, α_j and β are used, while the ones based on the exceedance probabilities $P(W > j|\boldsymbol{x})$, $-\alpha_j$ and $-\beta$ are used. We calculate the probability of a favorable outcome for each arm as:

$$p_d = \Pr\left(Y \geq 0|\alpha_j, \beta\right) = \Pr\left(W \geq 6|\alpha_j, \beta\right) = \sum_{j=6}^{10} \pi_{jd}\left(\alpha_j, \beta\right), d = 1, 2.$$

Piecewise exponential model

Ordinal outcome W can be modeled as a TTE outcome censored at integers. Assume there is a variable T, $T > 0$ where the value of W is defined by the outcome T: $W = 1$ if $0 < T \leq 1$, W=2 if $1 < T \leq 2$, and so on. This connection between the ordinal data and a latently modeled time to event endpoint, T, allows the use of software for TTE models to be utilized for ordinal outcomes. Consider a Bayesian piecewise exponential (PE) model for the outcome W:

$$W \sim \text{PE}(\lambda_1, \ldots, \lambda_k, c_1, \ldots, c_{k-1}),$$

where $\lambda_1, \ldots, \lambda_k$ are baseline hazard rates and $c_1, \ldots, c_{k-1}$ refer to each time segment. The PE model is a special case of the Cox proportional hazard (PH) model with exponential survival time, which assumes that the hazard function is constant:

$$\lambda(W|\boldsymbol{x}) = \lambda_0(W)\exp(\gamma_d)$$

with weakly informative prior $\gamma_d \sim N\left(0, 1.82^2\right)$. Extended from the Cox PH model with exponential survival time, the PE model assumes that the baseline hazard is constant within each segment. The segmented baseline hazard rates for the control arm are:

$$\boldsymbol{\lambda_0}(W = t) = \begin{cases} \lambda_1, & 0 < t \leq c_1 \\ \lambda_2, & c_1 < t \leq c_2 \\ \vdots & \vdots \\ \lambda_k, & t > c_{k-1} \end{cases}$$

with noninformative priors on the hazard rates:

$$\lambda_j \sim \text{Gamma}\left(n_j, \frac{n_j}{\mu_j}\right),$$

where $n_j = 0.01, j = 1, 2, \ldots 10, \mu_j = \begin{cases} 1, & j = 1, 2, \ldots, 9, \\ 0.0001, & j = 10. \end{cases}$

For each time segment the conditional probability for each ordinal outcome can be derived from the hazard rate in the interval:

$$c_{jd} = 1 - \exp\left(-\lambda_{0j}\gamma_d T_j\right), j = 1, 2, \ldots, 10.$$

The probability of each ordinal category is a function of the conditional probabilities c_{jd}

$$\pi_{jd} = \frac{c_{jd}}{\left(1 - \sum_{i=1}^{j-1} c_{id}\right)}, j = 1, 2, \ldots, 10.$$

The probability of favorable outcome for each arm can be written:

$$p_d = \Pr(W \geq 6|\lambda_0, \gamma_d) = \sum_{j=6}^{10} \pi_{jd}\left(\lambda_{0j}, \gamma_d\right).$$

Connection between PO and PE

In the PE model, when the TTE outcome variable is censored to integers, the cumulative probability of no event is equivalent to the exceedance probability for an ordinal outcome; this is true because the analysis of ordinal data is invariant to the codes used to label the levels. For demonstration purposes, additional Figures 1 and 2 for the exceedance probability based on the PO model and the cumulative probability of no event based on the PE model is provided in Appendix II.

For the PO model, the underlying assumption is that the odds ratio of one treatment relative to another is constant across all cumulative responses, this property is written based on exceedance probability definitions:

$$\frac{\text{logit}\left\{P(W > j|\boldsymbol{x}_2)\right\}}{\text{logit}\left\{P(W > j|\boldsymbol{x}_1)\right\}} = \log\left\{\frac{P(W_i>j|\boldsymbol{x}_2)/P(W_i\leq j|\boldsymbol{x}_2)}{P(W_i>j|\boldsymbol{x}_1)/P(W_i\leq j|\boldsymbol{x}_1)}\right\} = -\left(\boldsymbol{x}_2^T - \boldsymbol{x}_1^T\right)\beta = -\beta,$$

where $j = 1, 2, \ldots J-1$ and $\boldsymbol{x}_2 = 1, \boldsymbol{x}_1 = 0$ refer to treatment and control.

The PE model is a special case of the Cox PH model with constant hazard such that the baseline hazard function is constant on each time segment, which satisfies the PH property:

$$\log\left\{\frac{\frac{\lambda(W|\boldsymbol{x}_2)}{\lambda_0(Y)}}{\frac{\lambda(W|\boldsymbol{x}_1)}{\lambda_0(Y)}}\right\} = \log\left\{\frac{\lambda(W|\boldsymbol{x}_2)}{\lambda(W|\boldsymbol{x}_1)}\right\} = \left(\boldsymbol{x}_2^T - \boldsymbol{x}_1^T\right)\gamma.$$

Under the PE model, the hazard rate is constant for each segment. The hazard ratio (γ) is the ratio of instantaneous rates at each integer; we could first convert the γ to the relative risk (RR)[16]:

$$\text{RR} = \frac{1 - e^{\gamma\ \ln(1-p)}}{p},$$

where p is the probability of a favorable outcome for the control group, and then convert from RR to odds ratio (OR)[17]:

$$\text{OR} = \frac{1 - p \times \text{RR}}{(1-p) \times \text{RR}}.$$

As mentioned before, the odds ratio here is the reciprocal of the conventional definition $\frac{(1-p)\times\text{RR}}{1-p\times\text{RR}}$ since the proportional odds model is defined based on exceedance probabilities. Therefore, there's a 1:1 correspondence between γ and OR:

$$\text{OR} = \frac{1 - p \times \left(\frac{1-e^{\gamma\ \ln(1-p)}}{p}\right)}{(1-p) \times \left(\frac{1-e^{\gamma\ \ln(1-p)}}{p}\right)} = \frac{e^{\gamma \ln(1-p)}}{\frac{1-p}{p} \times \left(1 - e^{\gamma\ \ln(1-p)}\right)}.$$

Therefore, $\gamma < 1$ corresponds to OR < 1.

2.3 | Trial simulations

The previous section provided a detailed description of the three Bayesian models (dichotomous, continuous, and ordinal) to analyze the same indicator $p_d, d = 1, 2$, the probability of a favorable outcome for each arm. A more thorough investigation of the relative properties of the Bayesian models can be done through repeated trial simulations. We used

TABLE 3 Virtual subject response scenarios for Bayesian designs ($n = 834$)

Ordinal score j	Null effect size (Ctrl.)	Very small effect size	Small effect size	Medium effect size	Large effect size
Exceedance Prob: $P(Y \geq j)$					
−5	1	1	1	1	1
−4	0.99	0.991	0.992	0.992	0.993
−3	0.935	0.942	0.943	0.948	0.952
−2	0.854	0.867	0.871	0.881	0.891
−1	0.662	0.69	0.698	0.719	0.74
0	0.492	0.528	0.539	0.567	0.595
1	0.3	0.338	0.351	0.381	0.415
2	0.156	0.188	0.198	0.226	0.257
3	0.065	0.085	0.092	0.112	0.136
4	0.014	0.022	0.025	0.034	0.045
Prob. of favorable outcome: $P(Y \geq 0)$					
p_1 (Ctrl.)	0.492	0.492	0.492	0.492	0.492
p_2 (Trt.)	0.492	0.528	0.539	0.567	0.595
$p_2 - p_1$	0	0.036	0.048	0.075	0.104

Abbreviations: Ctrl., control; Prob., probability; Trt., treatment.

data from a large phase III TBI trial (ProTECT-III)[2] with a total sample size of 834 subjects for the prospective Bayesian design and evaluated the operating characteristics and estimation efficiency for the probability of favorable outcome via simulations. ProTECT-III used a sliding dichotomy of the GOS-E, and method and results have been published previously.[12] Under the Bayesian framework, null and alternative hypotheses are defined as different scenarios under which we assess the performance of the simulated trials.[18] The null and alternative hypotheses are $H_0 : \delta = 0$ vs $H_1 : \delta > 0$, where δ is the difference between the true treatment effect, and different effect sizes of δ ("no," "very small," "small," "medium," and "large") were shown in Table 3. Ten thousand trials were simulated under each hypothetical scenario for each of the three analysis designs (dichotomous, continuous with continuity correction, and ordinal). We started with a set of simulated trials with a fixed sample size of 834 subjects as the prospective design of phase III TBI trials, and then repeated the same procedure for the prospective design of phase II TBI trials with a smaller sample size of either 200 or 80 subjects based on the data simulation mechanism specified in Tables 4 and 5 of Appendix.

We used fixed and adaptive clinical trial simulator 6.4 (FACTS)[19] for the dichotomous and continuous analysis methods and approximated the ordinal design by the FACTS TTE design using a piecewise exponential model. As a comparison to the ordinal approximation in FACTS, we also used an interface from R[20] to JAGS and the "rjags" package[21,22] to write R programs for the ordinal design using the PO model, and the details are discussed in the following sections. This comparison is important because FACTS uses the TTE approximation of the ordinal design while JAGS uses an exact form of the ordinal design. Also, FACTS requires a licensing purchase while JAGS uses an interface from the free programming software R.

2.3.1 | FACTS

The primary endpoint was specified as dichotomous, continuous, or ordinal with different data simulation frameworks. Dichotomous data was directly simulated in FACTS:

1. Specify probability of favorable outcome $p_1 = \mathrm{P}(Y_1 \geq 0) = \mathrm{P}(Z_1 = 1)$ for the control arm.
2. Specify probability of favorable outcome $p_2 = \mathrm{P}(Y_2 \geq 0) = \mathrm{P}(Z_2 = 1)$ for the treatment arm.
3. Investigate different hypothetical effect size scenarios and run $I = 10\,000$ simulations for each.

Continuous data was simulated by bootstrapping from the known ordinal distribution:

1. Obtain the true ordinal probabilities for the control arm ($\boldsymbol{p}_1$) based on the original placebo data from the ProTECT-III trial.
2. Calculate the true ordinal probabilities for the treatment arm ($\boldsymbol{p}_2$) based on $\boldsymbol{p}_1$ and prespecified effect size under each hypothetical scenario (null or alternative).
3. Generate $N = n_1 + n_2$ ordinal subject responses (n_1 control, n_2 treatment) according to $\boldsymbol{p}_1, \boldsymbol{p}_2$ under each hypothetical scenario as external datasets.
4. Load the external datasets into the FACTS.

Sample a dataset of N subjects with replacement from the external dataset under each hypothetical scenario and repeat $I = 10\,000$ times.

Ordinal data was simulated directly in FACTS:

1. Specify a piecewise exponential hazard rate for subjects in the control arm.
2. Specify the hazard ratio for the treatment arm under different effect size scenarios.
3. Investigate different hypothetical effect size scenarios and run $I = 10\,000$ simulations for each scenario.

As for the execution aspects of the simulated trials, the accrual rate was simulated in FACTS using a mean of five subjects per week with no dropouts, according to the original trial property. Patients were randomized to two arms (treatment, control) with equal allocation (1:1) and their scheduled visit was 26 weeks after randomization. No interim analysis or response-adaptive-randomization (RAR) were considered since adaptation is not the focus of our study.

2.3.2 | JAGS

Similar to the mechanism of data simulation in FACTS, ordinal data was simulated based on parametric sampling under different effect sizes. Ten thousand datasets of ordinal data were directly used for the ordinal and continuous design and were dichotomized for the dichotomous design.

We applied Markov chain Monte Carlo (MCMC) methods for the Bayesian ordinal design and analysis using program R 4.1.3 and the package rjags[21,22] that calls on program JAGS 4.3.0, and figures were produced using the package ggplot2.[23] The JAGS code is provided as Appendix I in the Supplementary file. To be consistent with the settings in FACTS, JAGS models were fitted using 1000 burn-in draws of MCMC simulations, followed by 3 chains for 2500 draws with no thinning for statistical inference.

2.4 | Assessment

2.4.1 | Operating characteristics

The operating characteristics including type I error rate and power were evaluated.[24]

To make designs comparable, success boundaries were chosen to achieve similar type I error rates across designs via simulations through a model calibration process. Note that during this model calibration process, the more simulations we run for each scenario, the more precise the type I error will be, and the maximum 95% margin of error is $1.96\sqrt{0.5 * 0.5/10{,}000} = 0.003$ for 10 000 simulations. In our study, the stopping boundary for success was adjusted to ensure the desired one-sided type I error of 2.5%. Figures 3 to 5 in Appendix II plot the type I error rate by threshold based on FACTS simulations for the PE model used for the FACTS ordinal design with different sample sizes ($n = 834$, 200, 80). As the threshold increases, the proportion of simulations meeting the success criterion decreases. The type I error rate is controlled at the nominal level when the threshold is chosen as 0.9464 for the phase III design with sample size 834, and 0.9484 or 0.9436 for the phase II design with sample size 200 or 80.

2.4.2 | Estimation evaluation

The same indicator $p_d, d = 1, 2$ is the true probability of favorable outcome for each arm.

Let $\hat{p}_d$ be the posterior mean of p_d. The expected posterior probability of favorable outcome can be obtained as the average of $\hat{p}_d$ across all the $I = 10\,000$ simulations.

$$E(\hat{p}_d) = \frac{\left(\sum_{i=1}^{10\,000} \hat{p}_{di}\right)}{10\,000}.$$

The bias is used to evaluate the accuracy of the estimates:

$$\text{bias} = E(\hat{p}_d) - p_d.$$

The sample variance is used to evaluate the precision of the estimates:

$$\text{Var}(\hat{p}_d) = \frac{\sum_{i=1}^{10\,000} \left(E(\hat{p}_d) - \hat{p}_{di}\right)^2}{10\,000 - 1}.$$

We also looked at the MSE as a comprehensive measure of both accuracy and precision:

$$\text{MSE} = \text{Bias}^2 + \text{Var}(\hat{p}_d) = E\left[(\hat{p}_d) - p_d\right]^2.$$

3 | SIMULATION RESULTS

In this section, we report the simulation results comparing the three models under null and several alternative scenarios described in Section 2.3.

3.1 | Type I error and power

As mentioned before, the null scenario is the case with an effect size of zero, and the alternative scenarios are the cases where treatment is superior to control with an effect size greater than zero (very small, small, medium, or large). For all the Bayesian designs, the stopping boundary for success was adjusted to ensure the desired one-sided type I error of 2.5%, so we could compare power among all the Bayesian design candidates across models using simulations in FACTS or JAGS. The operating characteristics including type I error rate and power for the Bayesian designs of different sample sizes are presented in Figures 1 to 3.

Figure 1 shows the simulation results for the prospective phase III trial designs with a fixed sample size of 834 subjects under the different scenarios described in Table 3, where the x-axis is the effect size, and the y-axis denotes power. The FACTS results provided in Figure 1A show that the ordinal design has higher power compared to the dichotomous or continuous design with continuity correction across all the alternative scenarios. The JAGS results provided in Figure 1B show that the continuous design with continuity correction has higher power compared to the dichotomous or ordinal design across all the alternative scenarios. The detailed magnitudes of Figure 1 are also provided in additional Table 3 in Appendix III.

Figures 2 and 3 shows the simulation results for the prospective Bayesian phase II trial designs with a fixed sample size of 200 or 80 subjects, respectively, under the different scenarios described in Tables 4 and 5 of Appendix. Figure 2 shows that the ordinal design has higher power compared to the dichotomous or continuous design with continuity correction across all the alternative scenarios. FACTS results provided in Figure 3A show that the ordinal design has higher power compared to the dichotomous or continuous design with continuity correction across all the alternative scenarios. The JAGS results provided in Figure 3B show that the continuous design with continuity correction has higher power compared to the dichotomous or ordinal design under the small or large effect size scenarios, while the ordinal design outperforms the others under the medium effect size scenario. The detailed magnitudes of Figures 2 and 3 are also provided in additional Table 4 and 5 in Appendix III.

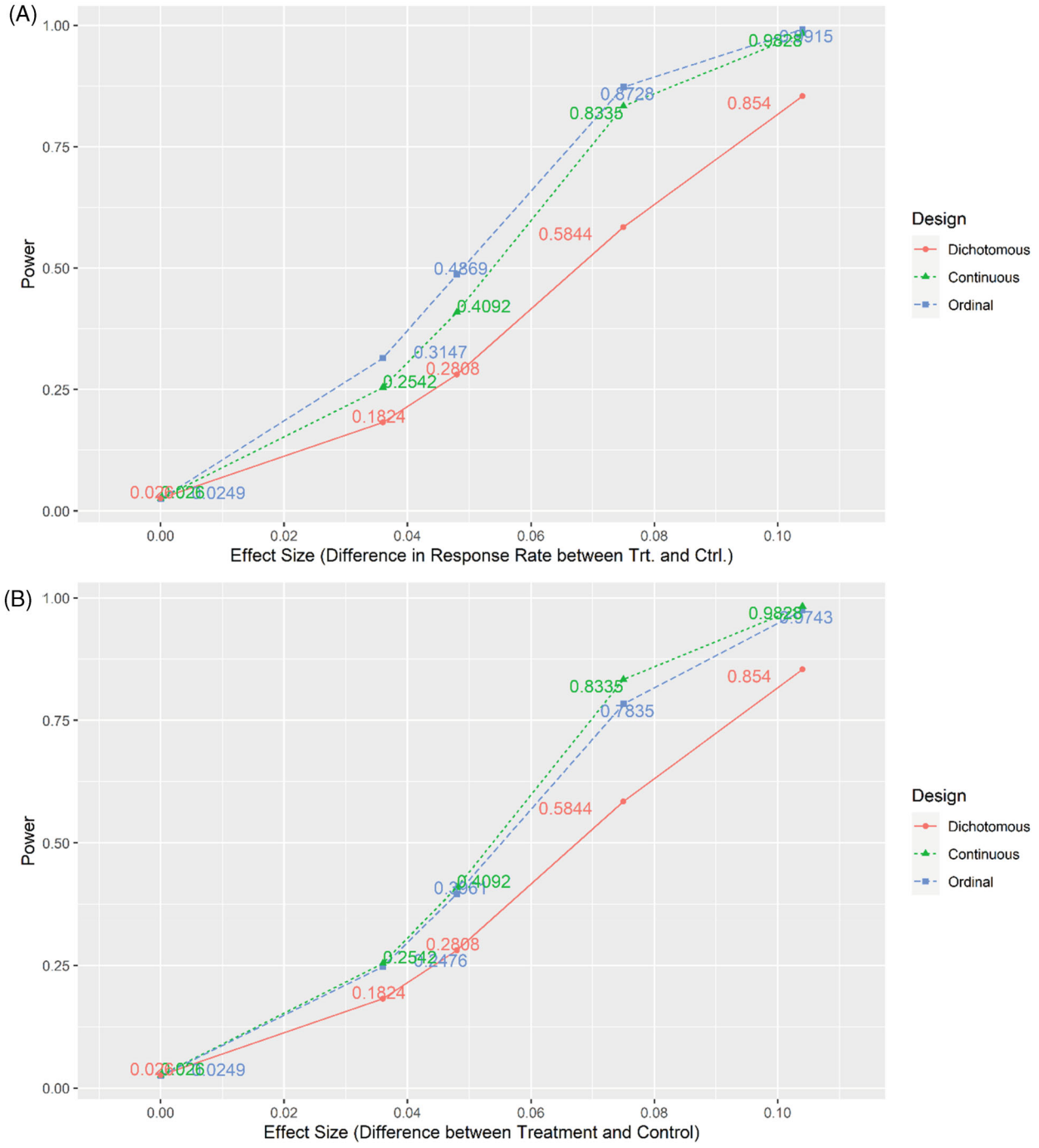


FIGURE 1 Compare power under different scenarios for Bayesian designs ($n = 834$). (A) uses the simulation results from FACTS for ordinal design, and (B) uses simulated results from JAGS for ordinal design

In summary, the power of the ordinal design is comparable to that of the continuous design and always higher than that of the dichotomous design.

3.2 | Bias, variance, and MSE

Figure 4 shows bias, variance, and MSE under the null and alternative scenarios described in Table 3 for a phase III trial simulated with a fixed sample size of 834 subjects. FACTS results provided in Figure 4A-C show that the dichotomous

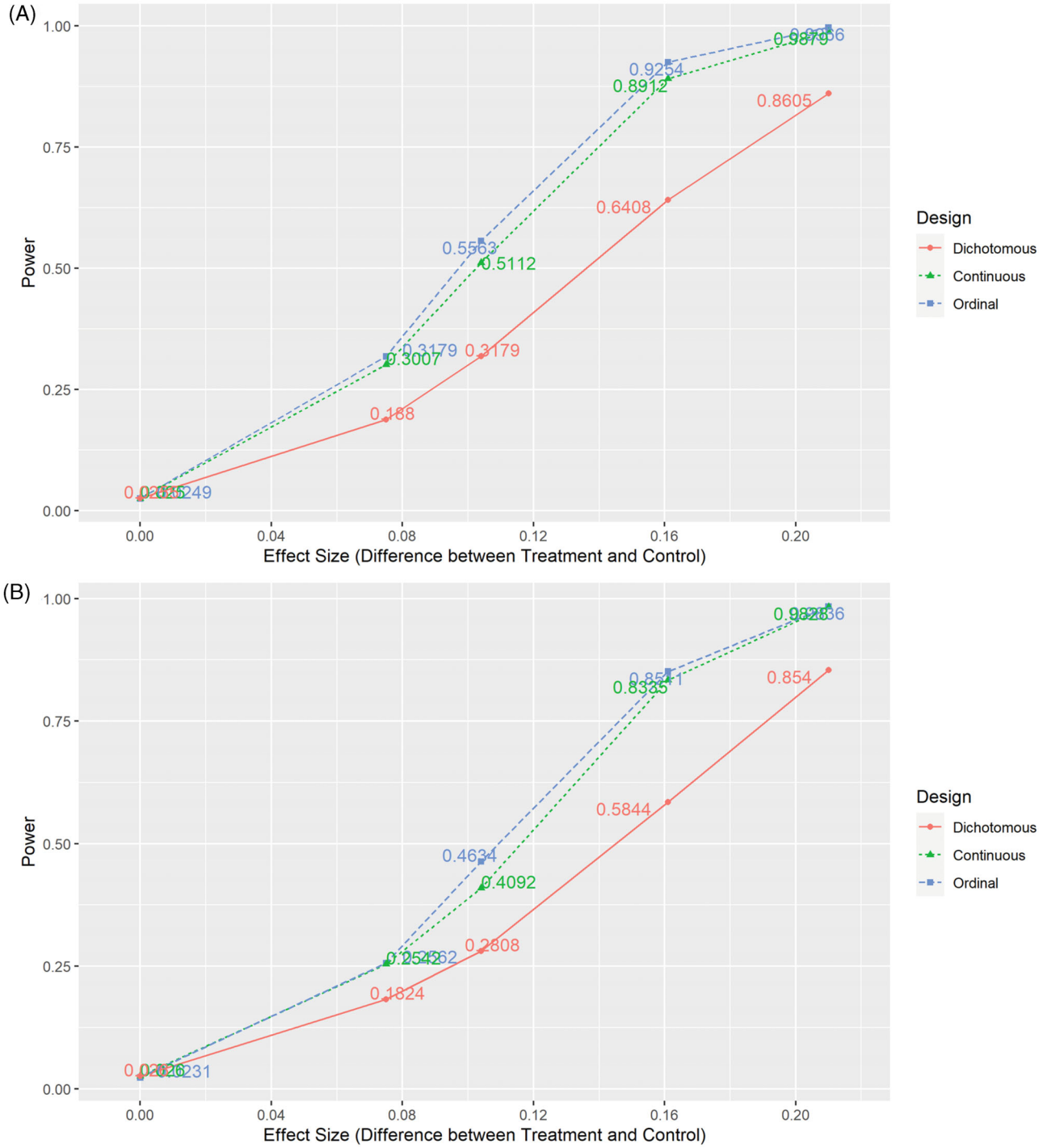


FIGURE 2 Compare power under different scenarios for Bayesian designs ($n = 200$). (A) uses the simulation results from FACTS for ordinal design, and (B) uses simulated results from JAGS for ordinal design

design produces unbiased estimates but has the highest variance, and the ordinal design produces the most biased estimates but has the lowest variance. The continuous design offers the lowest MSE. JAGS results are provided in Figure 4D-F show that accuracy of the ordinal design is similar to the continuous design but with slightly higher variance, and again the continuous design has overall lowest MSE.

Figures 5 and 6 show bias, variance, and MSE under the null and alternative scenarios described in Tables 4 and 5 of Appendix, for phase II trial design with a fixed sample size of 200 or 80 subjects, respectively. Both showed that the

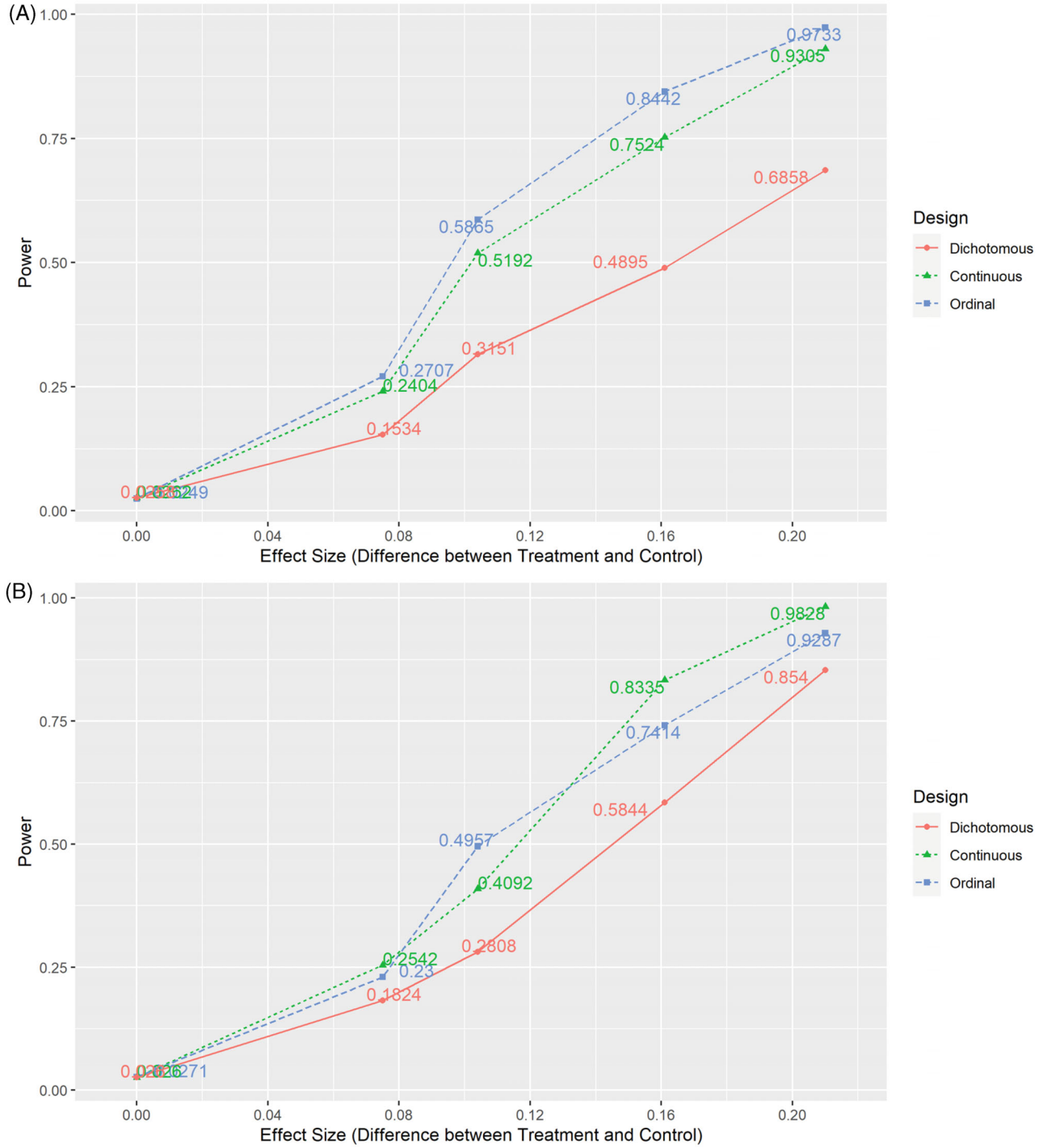


FIGURE 3 Compare power under different scenarios for Bayesian designs ($n = 80$). (A) uses the simulation results from FACTS for ordinal design, and (B) uses simulated results from JAGS for ordinal design

ordinal design produces estimates with similar variance but higher bias compared to the continuous design, and again the continuous design has overall lowest MSE.

As a summary for the operating characteristics of the trial designs using different analysis methods, the continuous design provides higher power compared to either the dichotomous or ordinal design when type I error is controlled at the nominal level. Overall, the continuous design has the lowest MSE. Therefore, the continuous design is the optimal choice as it has the highest power and lowest MSE, which indicates a combination of good accuracy and precision.

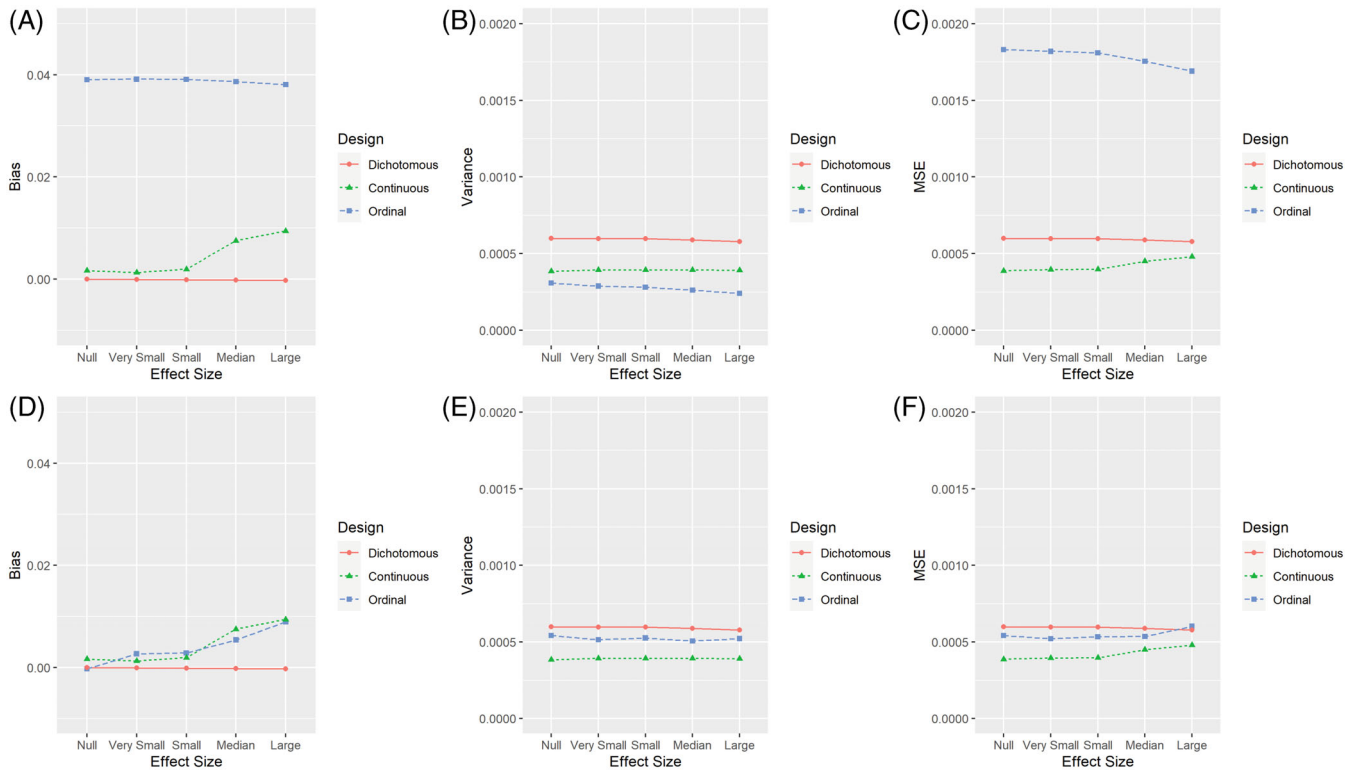


FIGURE 4 Compare bias, variance, and MSE under different scenarios for Bayesian designs ($n = 834$). (A to C) use the results from FACTS for ordinal design, and (D to F) uses the results from JAGS for ordinal design

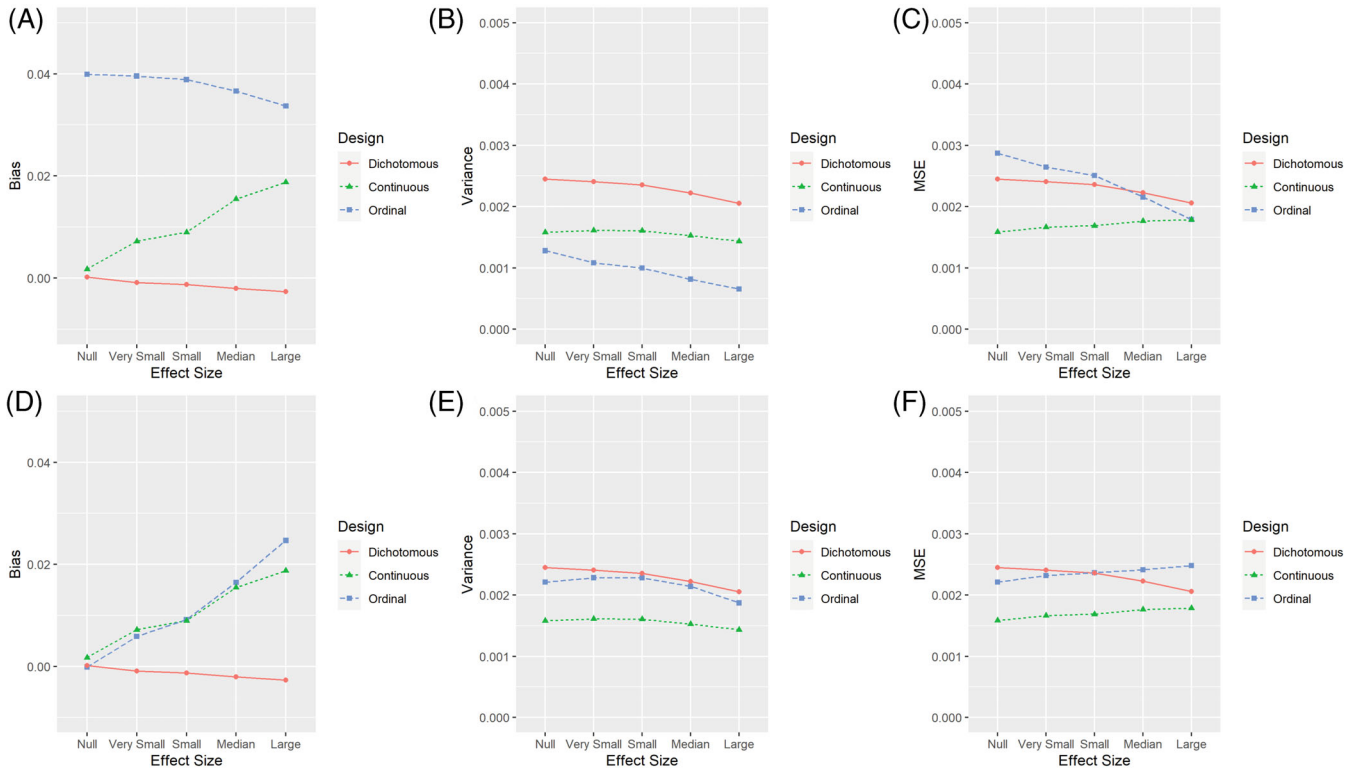


FIGURE 5 Compare bias, variance, and MSE under different scenarios for Bayesian designs ($n = 200$). (A to C) use the results from FACTS for ordinal design, and (D to F) uses the results from JAGS for ordinal design

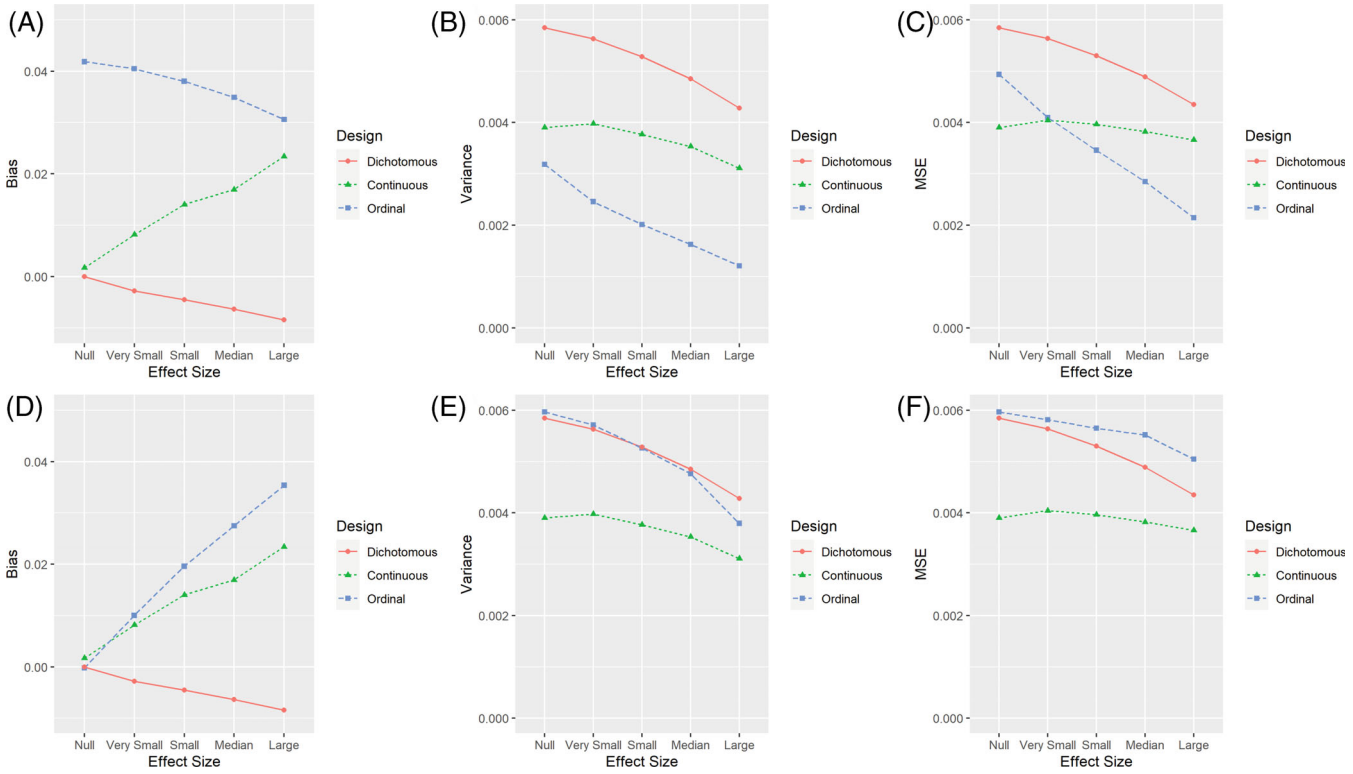


FIGURE 6 Compare bias, variance, and MSE under different scenarios for Bayesian designs ($n = 80$). (A to C) use the results from FACTS for ordinal design, and (D to F) uses the results from JAGS for ordinal design

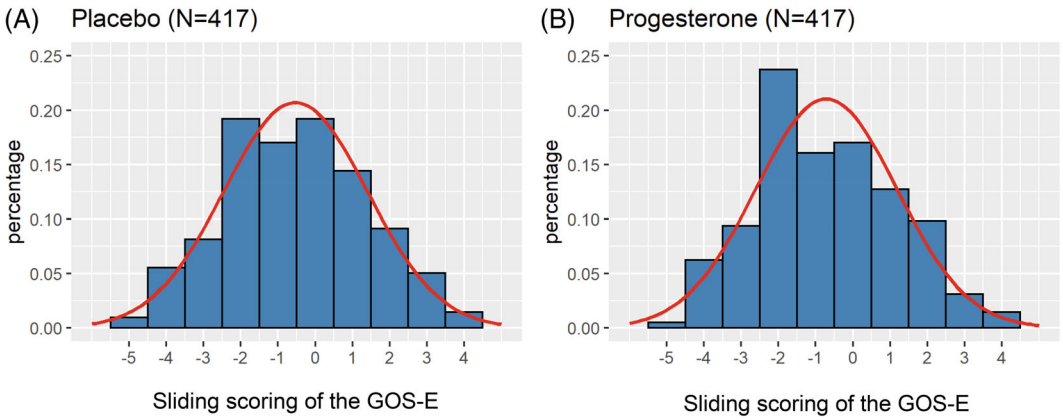


FIGURE 7 Histogram of ProTECT-III data with normal approximation

4 | APPLICATION TO PROTECT-III DATA

To further our findings, we applied these Bayesian models to the completed ProTECT-III trial. Figure 7 shows the histogram of the sliding scoring of the GOS-E for each of the two treatment arms. The normal approximation curve is imposed, where the mean and SD (SD) were calculated based on the observed data.

We applied the described Bayesian models to the ProTECT-III trial to evaluate Bayesian estimates of the same indicator, which is the probability of favorable outcome. To get accurate results for the re-analysis, JAGS models were fitted using 10 000 burn-in draws of MCMC simulations, followed by three chains for 10 000 draws with no thinning for statistical inference. We started the re-analysis of ProTECT-III trial data with a sample size of 834 subjects, and then repeated the same analysis for random samples of ProTECT-III trial data of either 200 or 80 subjects.

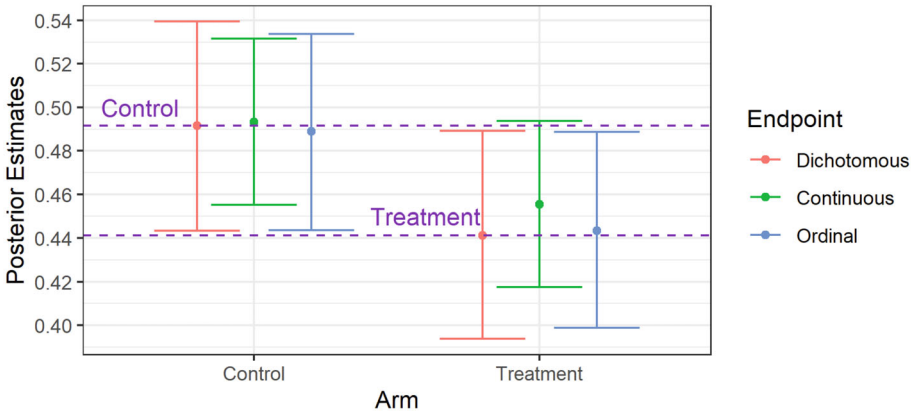


FIGURE 8 Bayesian analysis models applied to ProTECT-III data ($n = 834$). Posterior mean with 95% credible interval were presented

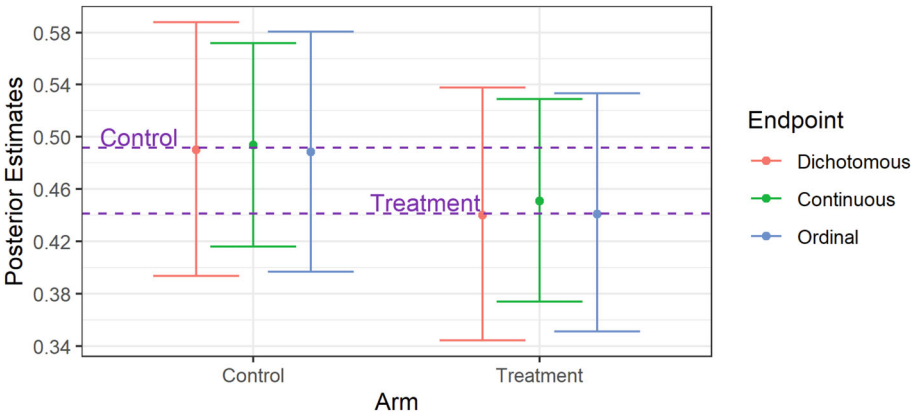


FIGURE 9 Bayesian analysis models applied to a random sample of ProTECT-III data ($n = 200$). Posterior mean with 95% credible interval were presented

Forest plots of the posterior means and Bayesian 95% credible intervals for probability of favorable outcome were shown in Figures 8 to 10, where the dashed purple lines serve as references and denote the observed proportion of favorable outcome for each of the two arms; red error bars denote the results using the dichotomous model, while green and blue represent the continuous and ordinal model respectively. In Figure 8, for the analysis of all the 834 subjects from ProTECT-III, the posterior estimates are a good fit across all the three designs as they overlap quite well with the reference lines, and continuous > ordinal > dichotomous in terms of precision based on the comparison of the widths of the 95% credible intervals. In Figures 9 and 10, the analysis of a random sample of 200 or 80 subjects from the ProTECT-III trial data, the story is consistent with the whole sample analysis results in Figure 8.

Some may be concerned about how the sliding thresholds are applied to a particular population. For example, some clinical trial populations may need calibration of the risk model (see for example Reference 25). Therefore, the actual risk may not be correct if one uses the risk score to create subgroups (with the four thresholds in Tables 1 and 2). Therefore, one may wonder how much those influence the results and comparison of methods.

To investigate this, we replicated the described models using the GOS-E raw score, ranging from 1 to 8 (death to upper good recovery). For the dichotomous model, we applied the fixed threshold for the favorable outcome (≥ 5) referenced in the introduction. The continuous and ordinal models used the GOS-E raw score. The IMPACT core model probability of poor outcome was mean-centered and included as an adjustment in each of the models. We then compared the three models for supplying inference regarding the probability of a favorable outcome, adjusting for the baseline risk. The parameter associated with the baseline risk variable had a non-informative prior.

These models yield similar results to the full sliding analysis. The continuous > ordinal > dichotomous is preserved in terms of precision based on the comparison of the widths of the 95% credible intervals. The widths for the 95% credible

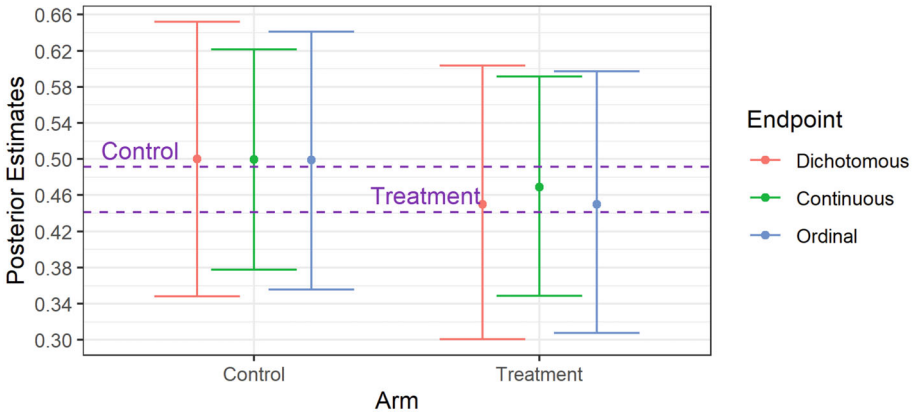


FIGURE 10 Bayesian analysis models applied to a random sample of ProTECT-III data ($n = 80$). Posterior mean with 95% credible interval were presented

intervals were approximately 0.10, 0.08, and 0.09 for the dichotomous, continuous, and ordinal, respectively, for both treatment arms in Protect III. We also replicated the analysis taking out the adjustment for baseline risk and performed the analysis and we obtained widths of 0.09, 0.08, and 0.08 for the dichotomous, continuous, and ordinal, respectively. This resulted in the continuous = ordinal > dichotomous in terms of precision based on the comparison of the widths of the 95% credible intervals. This result supports the use of a continuous model should other analysts choose to stay clear of the sliding scoring and want to do within-trial risk adjustment or no risk adjustment, the relative precision of the three methods remains unchanged.

5 | DISCUSSION

5.1 | Innovations

Previous analyzes for TBI clinical trials treated the endpoint as either dichotomous using a generalized linear model based on a sliding dichotomy[2] or, as exploratory use, continuous using a linear model without continuity correction based on a sliding scoring of the GOS-E.[12] No studies utilized the ordinal nature of the GOS-E for estimation of the probability of favorable outcome in TBI clinical trials. We filled this gap by comparing the ordinal design to the continuous and dichotomous design through a simulation study to evaluate power and estimation efficiency and apply these methods to a completed TBI clinical trial (ProTECT-III). To make all the designs comparable, one-sided type I error rates were controlled at the nominal level of 2.5%, and the same estimand—the binary indicator of achieving a favorable outcome—was derived in all designs.

To adjust for the fact that a continuous random variable was used to approximate a discrete variable when the data is non-normal, the normal distribution with continuity correction was considered for the continuous design. Based on simulation results, we found that when the data are non-normal, the continuous design with continuity correction had higher power and lower MSE compared to the dichotomous or ordinal design.

Another innovation of our study is to analyze the ordinal outcome using a piecewise exponential model and compare the results to a conventional ordinal model such as the PO model. The TTE outcome has never been repurposed to approximate an ordinal outcome. We compared the ordinal design using either the TTE approach in FACTS or the PO approach in JAGS to continuous or dichotomous design. For a fixed Bayesian design of 834 subjects with no interim analysis or allocation adaptation, it took 5 min in FACTS to simulate 10 000 virtual trials, but 2000 min in JAGS with parallel computing. We have shown that using the FACTS TTE design as an approximation of the ordinal design is accurate in terms of power calculation, given appropriate type I error control that is comparable to other designs. However, there could be potential bias caused by the rounding up of the exposure time due to the restriction of observations at visit times. In our simulation scenarios, the small percentage for some of the ordinal categories caused the ratio of the number of events to the number of exposure sensitive to the rounding up error. Therefore, we advise not using the FACTS TTE design for getting accurate estimates but only for the purpose of power estimation given strict type I error control.

5.2 | Limitations

A limitation to sliding dichotomy or the sliding scoring is that they only provides relative improvements and may result in a slight loss of statistical power. For sliding dichotomy, an intervention that improves outcome from vegetative to severe disability status in a subset of the patient population, may not have the same impact on a more general population. For sliding scoring, a change of three points could mean to either remain severely disabled or regain all functions.

We also need to be aware of the ordinal nature of the sliding scoring of the GOS-E when treating it as continuous. We cannot assume that the increment from "1" to "2" is the same as the increment from "2" to "3" without justifications. When continuous models based on the normal approximation are applied to ordinal data, it is implicitly assumed that there are equal distances between the discrete response levels. Without proper assumptions, applying continuous models to ordinal data can lead to misinterpretations of the data.[26] In the context of the GOS-E, the inherent assumption is that each improvement from one level of the GOS-E to the next is of equal clinical importance. For example, an effect size of 0.5 may reach statistical significance with a given sample size, but the change in clinical significance (less than 1 level) on the scale may not be obvious. Therefore, caution needs to be exercised in interpretating the continuous modeling of the GOS-E. Meanwhile, the fixed and sliding dichotomies demonstrate a clear advantage here as effect size is interpretable regardless of its magnitude, therefore it is more intuitive and clinically appealing.

To adjust for unequal distances, alternative weightings for the sliding score of the GOS-E could be considered, for example the utility-weighted modified Rankin scale (UW-mRS), which is an ordinal measure often used for evaluating neurologic disability, that is, the degree of disability or dependence in everyday life. Recently, the UW-mRS has been proposed as a new patient-centered primary outcome measure in stroke trials for acute stroke interventions.[27,28] A 2020 pooled analysis of 22 946 acute stroke patients derived UW-mRS from the mRS score and health-related quality of life on the European Quality of Life 5-dimensional questionnaire (EQ-5D) and found that the UW-mRS provides efficiency gains as a smaller sample size is required to detect a treatment effect in acute stroke trials through use of continuous score. Future work could propose a similar utility-weighted measure of the GOS-E score and test its statistical properties and construct validity. Converting to a utility-weighted measure of the GOS-E score would require placing a value on the different outcome levels which could come from costs to the quality of life to patients suffering from severe TBI.

Another aspect to consider is the validity of the model assumptions. If the PO assumption is violated, the odds ratio from the PO model only represents a kind of average odds ratio, and the treatment effect is not constant over the ordinal scale under the case of clinical trial design and analysis, which is a limitation of the methodology.

Future studies could also compare the different statistical models when accommodating for more complex trial designs. For example, the motivating phase II trial HOBIT is a dose selection study,[13] so it is important to assess the probability of selecting the best dose via trial simulations in addition to power comparisons. We could also look into the sequential monitoring using interim analyzes and response-adaptive-randomization to achieve better randomization scheme and therefore improve power or shorten trial duration.

6 | CONCLUSION

We used the data from a completed TBI trial for prospective Bayesian designs using dichotomous, continuous, or ordinal models, and evaluated the trial operating characteristics and estimation efficiency via simulations. The simulation results showed that the continuous design with continuity correction has higher power and lower MSE compared to the dichotomous or ordinal design. From a clinical perspective, we eliminated the sliding dichotomy approach which treats all scores meeting favorable criteria the same as it is very inaccurate (highest MSE). Since the choice of analysis method for the GOS-E in future TBI trials should be guided by power,[5] we recommend that future TBI clinical trials wishing to estimate the probability of favorable outcome should analyze the sliding scoring of the GOS-E as continuous with continuity correction. Also, current trials should consider the continuous model as a secondary analysis.

We also illustrated a connection between ordinal data and TTE data censored to integers allowing use of FACTS that utilize TTE-based models. FACTS provides similar power but biased estimates when compared to another Bayesian software JAGS. FACTS is much more computationally efficient compared to JAGS. Therefore, we advise using FACTS TTE design only for the purpose of power estimation given strict type I error control as it tends to overestimate the endpoint.

ACKNOWLEDGEMENTS

The data that motivates this clinical trial design comes from grants from the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH), BOOST-3 Study U01NS099046, HOBIT Study U01NS095926. We also want to acknowledge the suggestions from Berry Consultants Statisticians Tom Parke and Mark Fitzgerald regarding the FACTS Time-to-Event Design. The authors are grateful to Berry Consultants President Scott Berry for his comments and suggestions on an earlier draft of the paper.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data used in this study were generated via simulation. The JAGS code for the three models is available in Appendix I. The FACTS files are available on reasonable request from the corresponding author (Yu Wang).

ORCID

Yu Wang https://orcid.org/0000-0001-8780-1342
Byron J. Gajewski https://orcid.org/0000-0001-7589-8247

REFERENCES

1. Johnson WD, Griswold DP. Traumatic brain injury: a global challenge. *Lancet Neurol*. 2017;16(12):949-950. doi:10.1016/s1474-4422(17)30362-9
2. Wright DW, Yeatts SD, Silbergleit R, et al. Very early Administration of Progesterone for acute traumatic brain injury. *N Engl J Med*. 2014;371(26):2457-2466. doi:10.1056/nejmoa1404304
3. Weir J, Steyerberg EW, Butcher I, et al. Does the extended Glasgow outcome scale add value to the conventional Glasgow outcome scale? *J Neurotrauma*. 2012;29(1):53-58. doi:10.1089/neu.2011.2137
4. Bath PMW, Lees KR, Schellinger PD, et al. Statistical analysis of the primary outcome in acute stroke trials. *Stroke*. 2012;43(4):1171-1178. doi:10.1161/strokeaha.111.641456
5. Alali AS, Vavrek D, Barber J, Dikmen S, Nathens AB, Temkin NR. Comparative study of outcome measures and analysis methods for traumatic brain injury trials. *J Neurotrauma*. 2015;32(8):581-589. doi:10.1089/neu.2014.3495
6. Murray GD, Barer D, Choi S, et al. Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy. *J Neurotrauma*. 2005;22(5):511-517. doi:10.1089/neu.2005.22.511
7. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332(7549):1080. doi:10.1136/bmj.332.7549.1080
8. Deyi BA, Kosinski AS, Snapinn SM. Power considerations when a continuous outcome variable is dichotomized. *J Biopharm Stat*. 1998;8(2):337-352. doi:10.1080/10543409808835243
9. Dijkland SA, Voormolen DC, Venema E, et al. Utility-weighted modified Rankin scale as primary outcome in stroke trials. *Stroke*. 2018;49(4):965-971. doi:10.1161/strokeaha.117.020194
10. Roozenbeek B, Lingsma HF, Perel P, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15(3):R127. doi:10.1186/cc10240
11. McHugh GS, Butcher I, Steyerberg EW, et al. 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. *Clin Trials*. 2010;7(1):44-57. doi:10.1177/1740774509356580
12. Sharon D, Yeatts RHM, Meurer W, et al. Sliding scoring of the Glasgow outcome scale-extended as primary outcome in traumatic brain injury trials. *J Neurotrauma*. 2020;37(24):2674-2679. doi:10.1089/neu.2019.6969
13. Gajewski BJ, Meinzer C, Berry SM, et al. Bayesian hierarchical EMAX model for dose-response in early phase efficacy clinical trials. *Stat Med*. 2019;38(17):3123-3138. doi:10.1002/sim.8167
14. Brain Oxygen Optimization in Severe TBI, Phase 3 (BOOST3). National Library of medicine. https://clinicaltrials.gov/ct2/show/NCT03754114
15. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;5(8):e165. doi:10.1371/journal.pmed.0050165
16. Shor E, Roelfs D, Vang ZM. The "Hispanic mortality paradox" revisited: meta-analysis and meta-regression of life-course differentials in Latin American and Caribbean immigrants' mortality. *Soc Sci Med*. 2017;186:20-33. doi:10.1016/j.socscimed.2017.05.049
17. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ*. 2014;348:f7450. doi:10.1136/bmj.f7450
18. Walley RJ, Grieve AP. Optimising the trade-off between type I and II error rates in the Bayesian context. *Pharm Stat*. 2021;20:710-720. doi:10.1002/pst.2102
19. *Fixed and Adaptive Clinical Trial Simulator (FACTS). Version 6.4*. Berry Consultants; 2021.

20. R: a language and environment for statistical computing. R foundation for statistical computing; 2021. https://www.R-project.org/
21. JAGS. A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling; 2003.
22. rjags: Bayesian Graphical Models using MCMC. Version R package version 4–10. 2019. https://CRAN.R-project.org/package=rjags
23. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
24. FDA. Guidance for industry: Adaptive designs for clinical trials of drugs and biologics. Accessed April 21, 2022 https://www.fda.gov/media/78495/download
25. Zheng J, Zheng Y, Hsu L. Re-calibrating pure risk integrating individual data from two-phase studies with external summary statistics. *Biometrics*. 2022;78(4):1515-1529. doi:10.1111/biom.13543
26. Liddell TM, Kruschke JK. Analyzing ordinal data with metric models: what could possibly go wrong? *J Exp Soc Psychol*. 2018;79:328-348. doi:10.1016/j.jesp.2018.08.009
27. Lees KR, Bath PMW, Schellinger PD, et al. Contemporary outcome measures in acute stroke research. *Stroke*. 2012;43(4):1163-1170. doi:10.1161/strokeaha.111.641423
28. Lees KR, Khatri P, Alexandrov AV, et al. Stroke treatment academic industry roundtable recommendations for individual data pooling analyses in stroke. *Stroke*. 2016;47(8):2154-2159. doi:10.1161/strokeaha.116.012966

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang Y, Yeatts SD, Martin RH, et al. Selection of a statistical analysis method for the Glasgow Outcome Scale-Extended endpoint for estimating the probability of favorable outcome in future severe TBI clinical trials. *Statistics in Medicine*. 2023;42(25):4582-4601. doi: 10.1002/sim.9877