Hyperbaric Oxygen Brain Injury Treatment (HOBIT) Trial: A Multi-factor Design with Response Adaptive Randomization and Longitudinal Modeling

Byron J. Gajewski,[1] Scott M. Berry,[2] William G. Barsan[3], Robert Silbergleit [3], William J. Meurer[3,4] , Renee Martin[5], Gaylan L. Rockswold[6, 7]

[1]Department of Biostatistics, University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, USA

[2]Berry Consultants, 4301 Westbank Drive, Bldg B, Suite 149, Austin, TX 78746, USA

[3] Department of Emergency Medicine, University of Michigan, 24 Frank L. Wright Dr. Suite H – 3100, Ann Arbor , MI 48106, USA

[4] Department of Neurology and Stroke Program, University of Michigan, and Michigan Center for Integrative Research in Critical Care (MCIRCC), University of Michigan, Ann Arbor , MI 48106, USA

[5] Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon Street Suite 303, MSC 835, Charleston, SC 29425, USA

[6]Hennepin County Medical Center, 701 Park Ave, Minneapolis, MN 55415. USA

[7] Department of Neurosurgery, University of Minnesota, Mayo Mail Code 96, 420 Delaware St., S.E., Minneapolis, MN 55455, USA

Author emails:
1. Byron J. Gajewski: bgajewski@kumc.edu
2. Scott M. Berry: scott@berryconsultants.net
3. William G. Barsan: wbarsan@umich.edu
4. Robert Silbergleit: robie@umich.edu
5. William J. Meurer: wmeurer@med.umich.edu
6. Renee L. Martin: hebertrl@musc.edu
7. Gaylan L. Rockswold: Gaylan.Rockswold@hcmed.org

Author Manuscript

Abstract

The goals of phase II clinical trials are to gain important information about the performance of novel treatments and decide whether to conduct a larger phase III trial. This can be complicated in cases when the phase II trial objective is to identify a novel treatment having *several* factors. Such multi-factor treatment scenarios can be explored using fixed sample size trials. However, the alternative design could be response adaptive randomization (RAR) with interim analyses, and additionally, longitudinal modeling whereby more data could be used in the estimation process. This combined approach allows a quicker and more responsive adaptation to early estimates of later endpoints. Such alternative clinical trial designs are potentially more powerful, faster, and smaller than fixed randomized designs. Such designs are particularly challenging, however, since phase II trials tend to be smaller than subsequent confirmatory phase III trials. The phase II trial may need to explore a large number of treatment variations to ensure that the efficacy of optimal clinical conditions is not overlooked. Adaptive trial designs need to be carefully evaluated to understand how they will perform and to take full advantage of their potential benefits. This manuscript discusses a Bayesian RAR design with a longitudinal model that uses a multi-factor approach for predicting phase III study success via the phase II data. The approach is based on an actual clinical trial design for the Hyperbaric Oxygen Brain Injury Treatment (HOBIT) trial. Specific details of the thought process and the models informing the trial design are provided.

Key words: Bayesian adaptive design; multiple factors; phase II clinical trial

3

4

Introduction

This manuscript describes a phase II clinical trial adaptive design for selecting the combination of hyperbaric oxygen (HBO2) treatment dose parameters (pressure, frequency, and intervening normobaric hyperoxia [NBH]) that provides the greatest improvement in the rate of good neurological outcome versus standard care for subjects with severe traumatic brain injury (TBI). A second goal of this phase II trial is to determine if there is any factor combination of HBO2 treatment that has at least a 50% probability of demonstrating improvement in the rate of good neurological outcome versus a control (i.e. standard care) in a subsequent phase III confirmatory trial, assuming to be 500 in the control and 500 in the novel arm.

Despite numerous clinical trials for treatment of TBI, subjects with TBI have high mortality and poor outcomes ([1]). Preclinical and clinical investigations indicate that HBO2 is physiologically active in reducing brain injury and improving outcomes in severe TBI ([2]). There are peer-reviewed published animal studies from well-established research laboratories which indicate that HBO2 potentially improves outcome from TBI by multiple mechanisms ([3]). By markedly increasing oxygen (O2) delivery to the traumatized brain, HBO2 can improve cellular energy metabolism, attenuate cell signaling and cytosolic ischemic cascades, and reduce subsequent necrotic and programmed cell death. Many clinical investigations in HBO2 have steadily corroborated that HBO2 in comparison with standard care significantly improves markers of oxidative metabolism in relatively uninjured brain as well as pericontusional tissue, reduces intracranial hypertension, demonstrates improvements in markers of cerebral toxicity and improves clinical outcome. However, important information, optimizing the HBO2

5

treatment factor combination in terms of pressure, frequency, and whether NBH delivered following the HBO2 treatment, is required prior to a definitive clinical efficacy study. Preclinical investigations working with TBI models have used pressures varying from 1.5 to 3.0 atmospheres absolute (ATA). Clinical investigators have used pressures varying from 1.5 to 2.5 ATA. However, the lungs in severe TBI subjects have frequently been compromised by direct lung injury and/or acquired ventilator pneumonia and are susceptible to O2 toxicity. Working within these constraints, it is essential to determine the most effective HBO2 dose schedule without producing O2 toxicity and clinical complications. This proposed clinical trial is designed to answer these questions and to provide important information for a confirmatory efficacy phase III trial.

The primary endpoint is the severity adjusted Glasgow Outcome Scale - Extended (GOS-E, binary response) at 6 months after the patient has enrolled in this study. Moreover, GOS-E at 30 days could be used for predicting 6-month GOS-E, allowing for accelerated learning of the primary endpoint through the longitudinal modeling. The combination of the three HBO2 treatment factors with different levels - pressure (1.5, 2.0 and 2.5 ATA), dose frequency (once a day [QD] and twice daily [BID]), and NBH (with and without) - will be studied. Not all possible combinations are used because of potential O2 toxicity, and the trial will explore the efficacy of nine different active arms in comparison to the control arm. If there is at least one experimental arm with sufficient probability of being better than control at phase II, the combination of that active arm will be selected for the future phase III trial to confirm the HBO2 efficacy. Another

design constraint is the sample size is about 200 enrolled subjects from approximately 15 clinical centers. The sample size is relatively small due to budget constraints.

Given these constraints, this multi-factor treatment design is challenging: there are three dimensions to study, the sample size is small, and the tool that provides the endpoint (GOS-E) is subject to variability from the heterogeneity of the disease. In addition, the added variability from posterior predictive distributions used for forecasting phase III success contributes noise.

Therefore it would be quite attractive to garner model efficiencies to improve later prediction. This manuscript, using a detailed phase II clinical trial design as example, illustrates a Bayesian adaptive design is *useful*, which has been tailored for the case of multiple factors for predicting phase III success using the phase II data. Bayesian predictive probabilities are *useful* to drive learning, which in itself is not necessarily novel for a phase II trial, but has been under-utilized. A potential model's efficiency can be improved with a main effects factor model based on the strong assumption that there are no interactions. We also use the main effects model in this design for getting information from the other arms. This could have other applications, as studying related treatments is common and could be done more efficiently if such treatments are not assumed to be independent. Every subject will still help in learning about all three factors and predicting phase III success. However, it is believed there is a possibility of interactions within the treatment factors, so it is decided to use a pairwise independent model for the primary analysis. This can be considered a hybrid design of both a pairwise independent and main effects models. The approach based on an actual emergency medicine clinical trial design is called the Hyperbaric Oxygen Brain Injury Treatment (HOBIT) trial.

7

1. **Design Choice**

   Possible choices for the design are fixed sample size trials or response adaptive

randomization (RAR) with interim analyses for possible success and/or futility. Other options

include RAR with longitudinal (RAR+L) modeling because it can increase efficiency of the trial,

particularly when the short term endpoint is relatively predictive of the long term endpoint and

some relationship (linear, quadratic, etc.) between the two is approximately correct ([4]; [5]; [6];

[7]).

   RAR may provide clinical trial designs that are more powerful, faster, and smaller than

fixed randomized designs. However, phase II trials tend to have a smaller sample size than their

subsequent confirmatory phase III trials and may require exploration of numerous treatment

options to identify the combination of treatment parameters most likely to improve clinical

outcome. Therefore, trial designs need to be carefully studied in order to take full advantage of

the RAR approach ([8]; [9]). For example, a multi-factor treatment clinical trial quickly reduces

to a sparse amount of human subjects per treatment combination. In other words, efficiency is

achieved by considering the information of adjacent dose cells as informative rather than

independent. The RAR in dose finding strategy provides a useful approach for optimal clinical

trial design in the case of a single factor with multiple doses ([10]); there is also literature on the

two-factor clinical trial design ([11]; [12]; [13]; [14]), however, there is little literature on trial

design and sample size computations using multi-factor (e.g. three-factor) designs having RAR.

   The trial will utilize RAR to favor the better performing experimental arms and possibly

early stopping for success or futility. RAR implies using a predefined algorithm for changing the

8

treatment allocation during the trial based on efficacy data, while leaving the clinical investigators blinded. In some situations RAR allows for substantially smaller sample sizes and provides better conclusions by favoring arms that are performing better, and slowing enrollment to the arms that are performing relatively poorly ([15]). Although unlikely in the phase II multi-treatment space, early termination of the study can allow more rapid development of a promising treatment, or, more commonly, efficient identification and rejection of less or not effective treatments that are unlikely to be beneficial if pursued further.

2. Methods

This section presents the details for the phase II HOBIT trial. The goal of this phase II trial is to identify the best HBO2 treatment for subjects with severe TBI, which would optimally combine three HOBIT treatment factors with different levels: pressure (1.5, 2.0 and 2.5 ATA), dose frequency (BID and QD), and NBH (with and without). Furthermore, this manuscript also tries to predict the phase III success based on the phase II data.

2.1 Treatment Arms

There are ten treatment arms in the trial:

Arm (a)	Pressure (ATA)	Frequency	NBH
1	0	0	Without
2	2.0	QD	Without
3	2.5	QD	Without
4	1.5	QD	With
5	2.0	QD	With
6	2.5	QD	With
7	1.5	BID	Without
8	2.0	BID	Without
9	2.5	BID	Without
10	1.5	BID	With

9

We label the control arm as $a = 1$, and the experimental arms as $a = 2, 3, 4, 5, 6, 7, 8, 9,$ and 10, respectively.

2.2 Primary Endpoint

The primary endpoint is the GOS-E assessed at 6 months after the subject is enrolled in the clinical trial. The GOS-E is a binary response variable with the value of success or failure. We label it as Y_6. Additionally, the GOS-E is available one month after enrollment which will be used to inform a longitudinal model to predict the 6 month outcome response. Similarly, we label it as it as Y_1.

2.3 Primary Analyses

In the first analysis we define the maximally effective treatment (a_{max}) as the treatment with the greatest effect. For each experimental arm, we calculate the posterior probability of being superior to control, $Pr(\theta_a - \theta_1 > 0)$, where θ_a is the six-month GOS-E response rate for experimental arm a, and θ_1 is that for control arm. As clinical data are analyzed by the pairwise independent model, the phase II trial will be stopped if one of the three following cases occurs:

1. Early Success: At each interim analysis, the trial may stop accrual for expected success if $Pr(\theta_a - \theta_1 > 0) > 0.975$ for a_{max}. There must be at least 150 total subjects enrolled before the trial may stop for success. If a success stopping rule is met, then a final analysis will be conducted after all currently enrolled subjects have been followed to their final endpoint.

2. Early Futility: At each interim analysis, the trial may stop accrual for early futility if $Pr(\theta_a - \theta_1 > 0) < 0.55$ for a_{max}, therefore all arms would meet this inequality for futility.

3. Final Success: At the final analysis, the trial will be considered a success if $Pr(\theta_a - \theta_1 > 0) > 0.94$ for any a, this inequality would need to occur only for one arm. Otherwise the Phase II trial will fail.

We now explain the second analysis which is using this phase II data for the prediction of Phase III success. A prediction of Phase III success is if the maximally effective treatment has a greater than 50% probability of HBO2 treatment demonstrating improvement versus control in a subsequent confirmatory trial with size 500 in the control and 500 in the novel treatment. A prediction of Phase III success is only calculated if $Pr(\theta_a - \theta_1 > 0) > 0.94$ for the arm a_{max}. Note that this calculation is made using the main effects model rather than the independent model above. This should increase efficiency in this prediction.

2.4 Analysis Population

The *Intent-to-treat* population (ITT) will be used to analyze the data. The ITT population will include all randomized subjects, and they will be assigned to different arms based on the randomization information regardless of the treatment received.
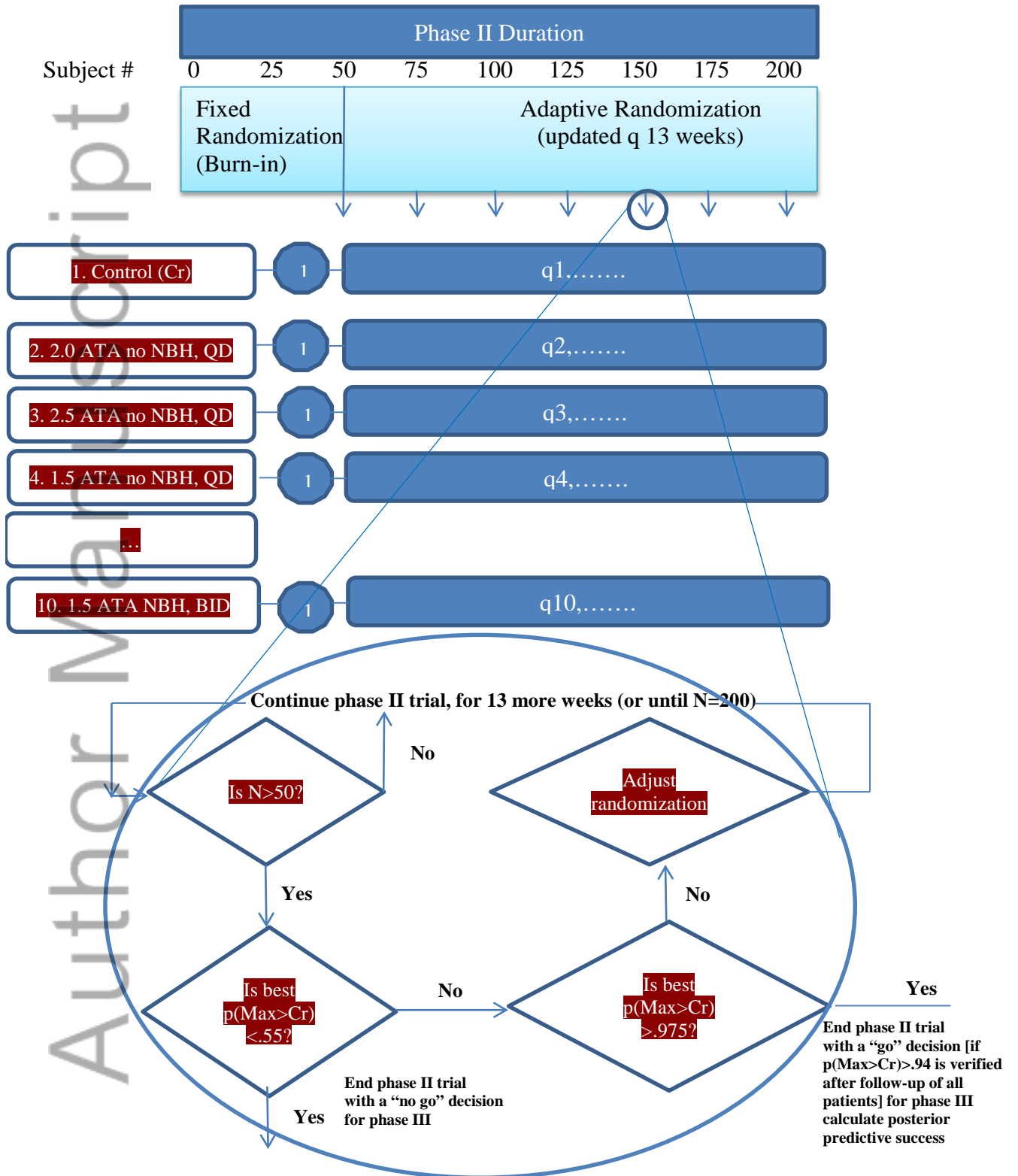
2.5 Adaptive Design

The purpose of this phase II adaptive clinical trial design is to explore the efficacy of different active arms in comparison to the control arm. The trial will not compare the active arms with each other. The trial will utilize RAR to favor the better performing experimental arms. If the efficacy within at least one experimental arm is promising enough, it will advance to a phase III trial and be compared for superiority to the control arm (Figure 1). The trial will use

11

both a pairwise independent and main effects multiple factor models to analyze the data based on the study objective, more detailed information will be described in section 2.7.

2.6 Randomization Introduction

1) <u>Burn-in Phase</u>: An initial burn-in period of 50 subjects is used in which these subjects are enrolled in a fixed randomization to the each arm. A ratio of 1:1:1:1:1:1:1:1:1:1 will be used for the burn-in period.

2) <u>Adaptive Randomization Phase</u>: After the initial burn-in period, adaptive randomization will be utilized. A vector of probabilities, $\boldsymbol{q}=(q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10})$, is created for randomizing to the experimental arms. A constant proportion of 20% of subjects will be enrolled to the control arm through the rest of phase II. Interim analysis will be executed quarterly to adjust the randomization probabilities based on the current interim analysis results. The probabilities will be proportionally set to each experimental arm based on arm with $a_{\max}$. More information and details is described in Section 2.7.5.

12

Figure 1: Study Design.

Phase II Duration

Subject # | 0 | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200

Fixed Randomization (Burn-in)

Adaptive Randomization (updated q 13 weeks)

1. Control (Cr) → 1 → q1,…….

2. 2.0 ATA no NBH, QD → 1 → q2,…….

3. 2.5 ATA no NBH, QD → 1 → q3,…….

4. 1.5 ATA no NBH, QD → 1 → q4,…….

…

10. 1.5 ATA NBH, BID → 1 → q10,…….

Continue phase II trial, for 13 more weeks (or until N=200)

Is N>50? — No

Adjust randomization — No

Yes

Is best p(Max>Cr) <.55? — No → Is best p(Max>Cr) >.975? — Yes

Yes

End phase II trial with a "no go" decision for phase III

End phase II trial with a "go" decision [if p(Max>Cr)>.94 is verified after follow-up of all patients] for phase III calculate posterior predictive success

13

2.7 Statistical Modeling

This section describes the statistical modeling used in the adaptive design and the primary analysis. The modeling is Bayesian in nature.

Two models- a *pairwise independent* response model and a *main effects* response model- are utilized in this study. All of the trial adaptations are driven by the *pairwise independent* model and we also identify the maximally effective experimental arm with this model. The *main effects* model will be used to predict phase III success, and this model assumes no interactions among the different level combinations of HOBIT treatment factors and will likely improve efficient prediction.

2.7.1 Pairwise Independent Response Model for Six-Month GOS-E Response

The primary outcome is six-month GOS-E response, and we label it for some subject *i* as $Y_{i,6}$. We model the six-month primary outcomes as Bernoulli distributed. The model is $[Y_{i,6}] \sim \text{Bernoulli}(\theta_{ai})$, where $a_i = 1,2,3,\ldots,10$ is the arm for subject *i*.

We label the probability of the six-month GOS-E response for arm *a* as θ_a. It is expected GOS-E response for control arm and novel arms have the following prior distributions: $\text{logit}(\theta_1) \sim N(-.41,.75^2)$, for the control arm; $\text{logit}(\theta_a) \sim N(0,1.75^2)$, for the experimental arms where $a=2,3,4,\ldots,10$. According to the previous clinical trials with the same endpoint, the control arm's prior on the response scale (θ_1) has a median of 0.40. If simulated data is fitted to a Beta distribution, the control arm's prior is equivalent to eight patients i.e. $\alpha_0 + \beta_0 \approx 8$, where α_0 and β_0 are Beta parameters. The novel arm's prior median on the response scale is higher than that of

14

the control arm at 0.50 but is much more vague since it is equivalent to only two patients (similar to a uniform distribution, $\alpha_0 + \beta_0 \approx 2$).

2.7.2 Main Effects Model for Six-Month GOS-E Response used for Phase III Prediction

The main effects model is $[Y_{i,6}] \sim \text{Bernoulli}(P_i)$, for subject i. We construct a main effects model for the GOS-E response rate that is a function of pressure, NBH, and duration. The logit transformation of P_i is modeled with a linear equation. By assuming no interaction among the main factors, this model has a lower number of parameters and is designed to increase confidence to predict phase III success. However, if there is an interaction, there would be less uncertainty in the predicted value but there would be bias. This scenario is explored later in trial simulations. The structure is

$$logit(\text{P}_i) = X_{i1}\mu + X_{i2}\alpha_{1.5ATA} + X_{i3}\alpha_{2.0ATA} + X_{i4}\alpha_{2.5ATA} + X_{i5}\gamma_{NBH} + X_{i6}\beta_{BID}$$

The Xs are 0 or 1 depending on the treatment factor combination assigned to subject i. μ represents the effect of control. The α's represent the additional effect of pressure relative to control. The γ's and β's represent the additional effect of NBH and BID, respectively. The main effects model relates to the control and experimental arms in the following way:

1. control μ

2. 2.0 ATA, without NBH, QD $\mu + \alpha_{2.0ATA}$

3. 2.5 ATA, without NBH, QD $\mu + \alpha_{2.5ATA}$

4. 1.5 ATA, with NBH, QD $\mu + \alpha_{1.5ATA} + \gamma_{NBH}$

5. 2.0 ATA, with NBH, QD $\mu + \alpha_{2.0ATA} + \gamma_{NBH}$

6. 2.5 ATA, with NBH, QD $\mu + \alpha_{2.5ATA} + \gamma_{NBH}$

15

7. 1.5 ATA, no NBH, BID $\mu + \alpha_{1.5ATA}$ $+ \beta_{BID}$

8. 2.0 ATA, no NBH, BID $\mu + \alpha_{2.0ATA}$ $+ \beta_{BID}$

9. 2.5 ATA, no NBH, BID $\mu + \alpha_{2.5ATA}$ $+ \beta_{BID}$

10. 1.5 ATA ,with NBH, BID $\mu + \alpha_{1.5ATA} + \gamma_{NBH}$ $+ \beta_{BID}$

The priors for GOS-E response for control arm and novel arm have the following prior

distributions: logit(μ)~N(-.41,.75^2), the control arm, and logit(*all other parameters*)~N(0,10^2).

The intuition regarding the control arm's prior is the same as the control's prior in the

independent model. The prior for the additional parameters' is essentially flat (i.e. uniform on the

real line).

2.7.3 Longitudinal Model

In addition to the final six-month endpoint, subjects will have a scheduled visit at 1

month. At any interim analysis, there may be subjects in each of the following categories:

subjects who have completed all the visits with known final endpoint; subjects who are still in

the visit process without final endpoint value; and subjects without data at all.

Let $Y_{i,6}$ be the final endpoint value for subject i and let $Y_{i,1}$ be the 1-month response. We

construct a longitudinal model to allow the unobserved 6-month endpoint to be imputed from the

partial data. The Beta Binomial longitudinal model updates two Beta distributions:

1. The posterior probability that a subject who is a responder at 1-month will be a responder

   at the 6 month endpoint:

$$Pr(Y_{i,6}=1| Y_{i,1} =1) \sim Beta(\alpha_1, \beta_1),$$

   where

16

$$\alpha_1 = 20 + |Y_{i,6} = 1, Y_{i,1} = 1|, \text{ and}$$

$$\beta_1 = 5 + |Y_{i,6=}0, Y_{i,1} = 1|.$$

2. The posterior probability that a subject who is a failure at 1-month will be a responder at the 6 month endpoint:

$$\Pr(Y_{i,6} = 1 | Y_{i,1} = 0) \sim \text{Beta}(\alpha_2, \beta_2),$$

where

$$\alpha_2 = 5 + |Y_{i,6} = 1, Y_{i,1} = 0|, \text{ and}$$

$$\beta_2 = 20 + |Y_{i,6} = 0, Y_{i,1} = 0|.$$

In the notation above, $|Y_{i,6} = 1, Y_{i,1} = 1|$ indicates a count of the number of subjects whose 6 month endpoint was observed to be a response and whose intermediate outcome at 1-month was also a response. The other formulas with absolute values have similar correspondence. We fit a set of models pooling data across all arms. Note that our priors are fairly diffuse, and each has a prior sample size equivalent to 25 subjects. This prior was informed from previous TBI studies.

2.7.4 Bayesian Quantities

The following Bayesian quantities used in this adaptive design are calculated at each interim analysis. From the joint posterior distribution the posterior probability that each arm, $a = 2,3,4,\ldots,10$ is the maximally effective arm, $P_a^{\max}$, is calculated. The arm with the largest $P_a^{\max}$ is called the *most likely maximum effective novel treatment.* The posterior mean and variance for each GOS-E response rate is also calculated. We label $V(\theta_a)$ as the posterior variance of the parameter θ_a. For GOS-E response rate the posterior probability that experimental arm is superior (larger response rate) to the control arm is calculated: $\Pr(\theta_a > \theta_1 | \text{data})$, where

17

a=2,3,4,…,10. Each of these Bayesian quantities are calculated using the data of all subjects who have already completed the trials or are still in visit process for each interim analysis.

2.7.5 Adaptive Randomization Details

The specification of the vector of probabilities for RAR is described in this section. The randomization vector is created by selection based on the posterior distribution of the GOS-E response for each arm.

The purpose of the adaptive randomization is to allocate subjects to the arms most likely to be maximum effective. In addition, the adaptive randomization could improve learning which is the maximum effective arm is in comparison with the control arm.

A *component,* labelled as V_a, is constructed for each arm. Set V_1=1, assuring 1/5 probability for control arm throughout the trial. The component for arms *a*=2,3,4,…,10 is

$V_a = 4P_a^{max}$ for *a*=2,3,4,…,10. The randomization vector, ***q***, is set as $q_a=V_a/5$ for *a*=1,2,3,4…,10.

2.7.6 Phase III Predictions

This adaptive design phase II clinical trial provides valuable information to predict the success probability of a phase III clinical trial which aims to confirm the efficacy and safety of optimal combination of HBO2 treatment factors for severe TBI in comparison with the standard care. The primary endpoint of the phase III trial will be the same as the one in phase II, i.e. sliding dichotomized GOS-E at 6 months, which will be primarily analyzed by a chi-square test in the phase III study. The sample size for phase III is 500 in the control and 500 in the novel arms, totally *n*=1000, based on *α*=.05 and 2-tailed.

Taking the maximally effective arm from phase II trial simulations, we calculate the posterior predictive probability there is a >50% probability of hyperbaric treatment demonstrating improvement in the rate of good neurological outcome versus standard treatment in a subsequent phase III confirmatory trial. This is calculated by the main effects model for the experimental arm that shows the maximum efficacy. To accomplish this the posterior predictive distribution for future responses of control and novel arms, we label y_1^P and y_a^P, for arm $a=2,3,...,10$, respectively, and they could be calculated via the following formula:

$$p(y_1^P, y_a^P \mid \mathbf{y}) = \int \binom{500}{y_1^P} P_1^{y_1^P}(1 - P_1)^{500 - y_1^P} \binom{500}{y_a^P} P_a^{y_a^P}(1 - P_a)^{500 - y_a^P} p(\mathbf{\theta} \mid \mathbf{y}) d\mathbf{\theta},$$

where $\mathbf{y}$ is the observed data from the phase II trial and $\mathbf{\theta}$ is a vector of parameters from the main effects model. This distribution is calculated using simulation. Suppose that a' is the best HBO2 experimental arm identified in phase II, then the appropriate predictive chi-square test statistic is $X_{a'}^{2P}$ which is calculated by y_1^P and $y_{a'}^P$. To achieve a > 50% success probability of phase III, two conditions must be satisfied: (1) the main effects model indicates a > 94% probability that treatment of arm a' is better than that in control arm; (2) there is a > 50% probability that $X_{a'}^{2P} > 3.841$.

2.8 A Comparison of Pairwise Independent and Main Effects Models

For the purposes of comparing the two models in a closed form situation, the primary endpoint should be considered as a continuous response. Using a flat prior for regression parameters and known variance, the posterior standard errors of the model parameters representing the parameter estimation could be easily calculated. Consider the burn-in period, it

19

can be easily shown, for example, that the posterior standard error of the pairwise independent

model is $\sigma\sqrt{\frac{1}{5}+\frac{1}{5}} = 0.6325\sigma$ when ATA is 2.0. The standard error within main effects model is

0.5578σ, thus the main effects standard error is 88% of the one of the pairwise independent

model, indicating the potential increase in efficiency of the phase III use of the main effects

model. More explicated information will be provided in the illustration of Table 5 in the results

section. Obviously, this reduction comes with added bias in case that an interaction does exist.

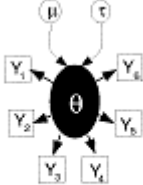**2.9 Software and Computations**

Computations were performed using three types of software: Fixed

and Adaptive Clinical Trial Simulator (FACTS) ([16]), R ([17]), and Windows Bayesian

inference Using Gibbs Sampling (WinBUGS) ([18]). General functions of these software are

specified in Table 1. The pairwise comparisons model with the longitudinal modeling and RAR

was performed in FACTS. The main effects model was performed using simulation in

R2WinBUGS with custom coding.

First, FACTS is a software used to rapidly design, compare, and simulate both fixed and

adaptive trial studies. It is built on compiled low-level languages, such as Fortran and C++, thus

it runs quite fast. Moreover, FACTS is flexible and easy to use since it could be accessed through

an interactive graphical user interface. However, currently, FACTS does not have the capability

to implement a main effects model. We decided to use FACTS to simulate the pairwise

independent model taking advantage of its flexibility and speed, then use the data output to call a

program that was written specifically for phase III predictions in R2WinBUGS. The posterior

20

simulated draws in FACTS were 1,000 burn-in and then 2,500 draws for inference. In

WinBUGS the burn-in was 1,000 and 1,000 draws for inference.

Table 1: Software we applied.Software	Function
	Simulate data and accrual pattern.
	Fit model for response and longitudinal modeling.
	Calculate all Bayesian quantities.
	Perform response adaptive randomization and allocate virtual patients to appropriate arms.
	Tally all trial operating characteristics of phase II.
	Pull data from FACTS into R.
	Call WinBUGS program.
	Tally all trial operating characteristics of phase III prediction.
	Fit main effects model and predict the posterior probability of phase III success.

2.9.1. Simulations

In FACTS^{TM} there is an option for defining dichotomous longitudinal response profiles that allows us to specify the overall transition probabilities between responder ('1') and non-responder ('0'). The transition probabilities method for generating longitudinal responses simulates the response observed at each visit by using the probability that a subjects becomes or remains a '1' from one visit to the next – and all subjects start with a response of 0.

We specify for each visit,

- the probability of a subject whose response was a '0' at the previous visit having a response of '1' at this visit

- the probability of a subject whose response was a '1' at the previous visit having a response of '1' at this visit

However these probabilities give a particular probability a subject has a response of '1' at the final visit, so they need to be modified for each arm in each dose response profile to give the desired final probability of response. This is done by numerically determining for each final response rate to simulate, a single value which when added to all the specified transition probabilities in the log-odds space yield the desired probability of final response.

With two visits (1-month and 6-months) and first visit probability of 0->1 of 0.5 then at the second visit probabilities of 0->1 of .2 and of 1->1 of 0.8, the probability of a final response is 0.5 (Table 2).

Table 2: Longitudinal data profile with initial transition probabilities.

Visit	Month	Prob 1->1	Prob 0->1
1	1	-	0.5
2	6	0.8	0.2

If a response profile calls for the probability of a final response to be simulated with a probability of 0.7, a fixed offset in log-odds is found which when applied to all the transition probabilities results in the desired final probability of a response. Replicating this by hand yields an offset of 0.68, which results in the adjustment in Table 3. All scenarios are adjusted in this fashion.

Table 3: Longitudinal data profile after adjustment.

Visit	Month	Prob 1->1	Prob 0->1
1	1	-	0.664
2	6	0.888	0.330

GOS-E responses at six-months of all the treatment arms within different case scenarios are presented in Table 4. The first case is referred to as the *null hypothesis* as each of the arms has identical GOS-E responses. All the novel arms have no improvement in terms of GOS-E response in comparison with the control arm. The remaining case scenarios explore the different GOS-E responses for the experimental arms including one case where harm is exhibited. The final case investigates the interactions of medium effect of pressure associated with NBH and frequency, which aims to comprehend the possible robustness of the main effects model.

Table 4: The adaptive clinical trial design evaluation via six Case Scenarios.

Case Scenario	1.0 Control	1.5 NBH	2.0, NBH	2.5, NBH	1.5, BID	2.0, BID	2.5, BID	1.5, NBH, BID	2.0	2.5
1. Null hypothesis	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
2. Small treatment effect	0.4	0.45	0.5	0.43	0.45	0.5	0.43	0.48	0.48	0.4
3. Medium treatment effect	0.4	0.5	0.55	0.48	0.5	0.5	0.48	0.55	0.5	0.43
4. Large treatment effect	0.4	0.57	0.7	0.52	0.57	0.7	0.52	0.65	0.63	0.45

23

5. Harmful treatment effect	0.4	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
6. Medium Interaction	0.4	0.4	0.55	0.4	0.55	0.40	0.55	0.4	0.4	0.4

3. Results

In this section we summarize the results of several simulation cases as well as one additional null scenario case to control type I error of the design. For each case, 1,000 trials are simulated. We present the results as a function of GOS-E response for each arm at six-month.

For all simulations in this section, we assume an accrual rate of 1.75 subjects per week. No drop outs are assumed.

The study is considered a success if a target duration arm is identified and phase III is recommended to be carried out. In the simulations, if a trial enters a possible success stage, the trial is stopped in the simulation.

We performed six sets of trial simulations based on the various cases of response to calculate the trial operating characteristics, i.e. power, futility probability, sample size, duration, and subject allocation, which are presented in Table 5. The first four cases range from no effect (null), to large treatment effect. All these four cases have no interaction between factors. We could clearly see an increase in power (starting with a 20% type I error rate) but a decrease in futility rates as the effect increases. Because the null trial has a higher chance to stop for futility, the balance switches in a higher probability to stop for success as the benefit moves to large. Both the sample size and duration increase from null hypothesis to medium treatment effect, then go down in large treatment effect. The percentage of patients placed in the best pressure (2.0 ATA) arms generally increases as treatment effect increases. The probability of phase III

24

success is uniformly larger than the power for phase II because of the added efficiency of the main effects model. The fifth case has explored characteristics, including the futility with a high rate, of the harmful treatment effect. Additionally, the medium interaction case has similar operating characteristics to the medium treatment effect case without interaction except that its probability (>50%) of phase III success prediction is lower than the phase II power. *Of note, the estimated sample size of fixed trials which is 325 and would give similar power is substantially larger.*

Table 5: Simulated trial operating characteristics.

Case	Power phase II	Futility Prob.	Size (n)	Duration (wks)	%Patient allocated to 2.0 ATA pressure arms	Probability (>50%) of phase III success*
1. Null hypothesis	0.20	0.34	176	118	33%	0.20
2. Small treatment effect	0.48	0.13	186	129	38%	0.51
3. Medium treatment effect	0.65	0.06	187	131	36%	0.71
4. Large treatment effect	0.96	0.01	174	125	45%	0.98
5. Harmful treatment effect	0.09	0.57	158	102	33%	0.08
6. Medium Interaction	0.63	0.09	185	106	33%	0.57

*New calculation based on main effects model (S=1000).

Figure 2 illustrates the patient allocation comparison of RAR+L and fixed one based on a single simulation's MCMC results. More subjects are placed on control for the RAR+L. There are also more subjects placed on the better experimental arms, and the top two are the third and sixth arms. Further the incremental improvement in the efficiency is shown by the decrease in the posterior standard deviation from pairwise independent model to a main effects model, illustrating the increase in prediction ability by the latter model. The standard deviation (Figure 3)

for the main effects model was on average 79% of the pairwise independent model, which is very similar to the approximate approach discussed in section 2.9.

Figure 2: Allocation of subjects across arms for one simulated medium effect case for RAR+L versus the fixed designs. Probability arm a=2,3,4,,…,10 is better than control.
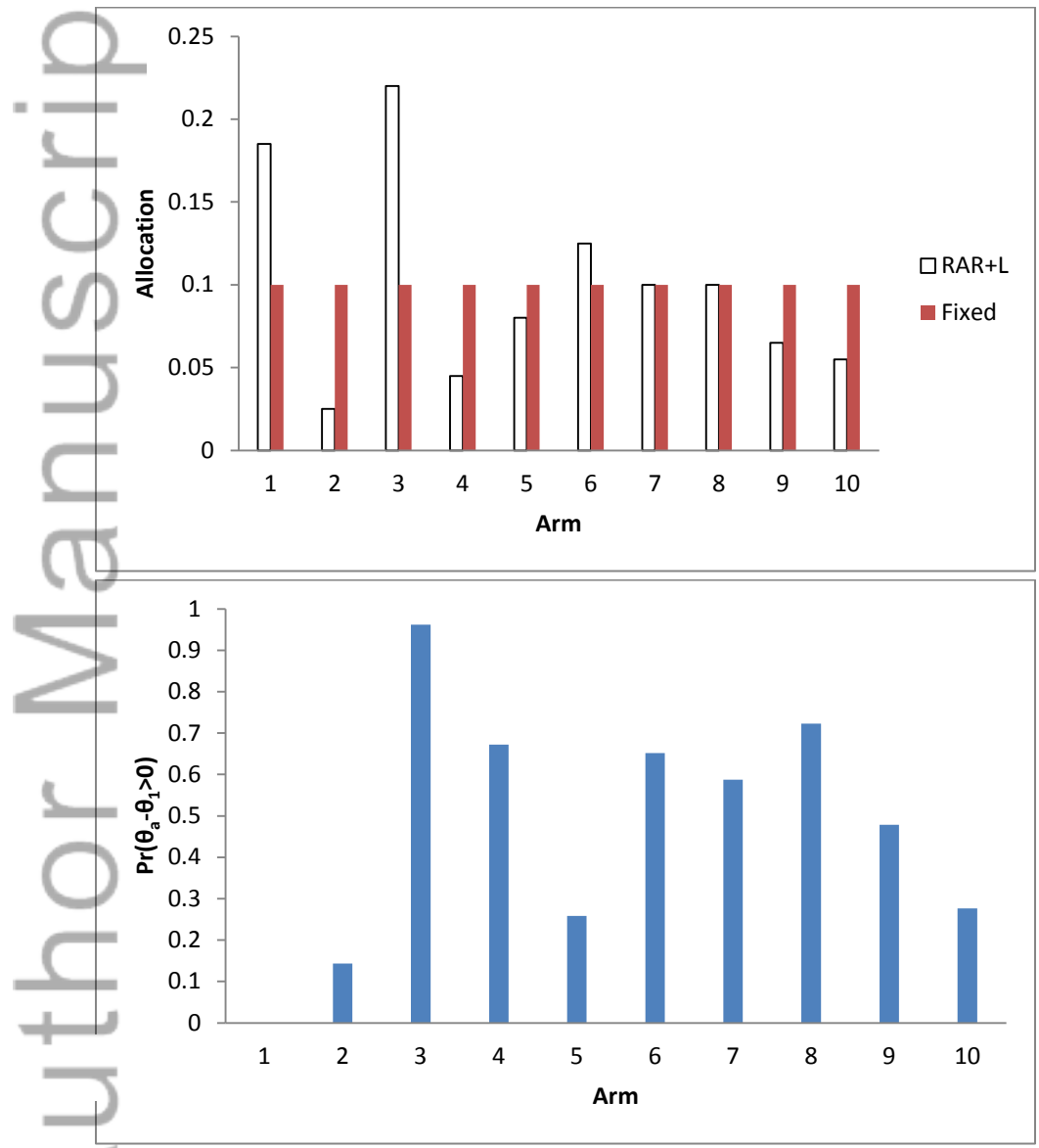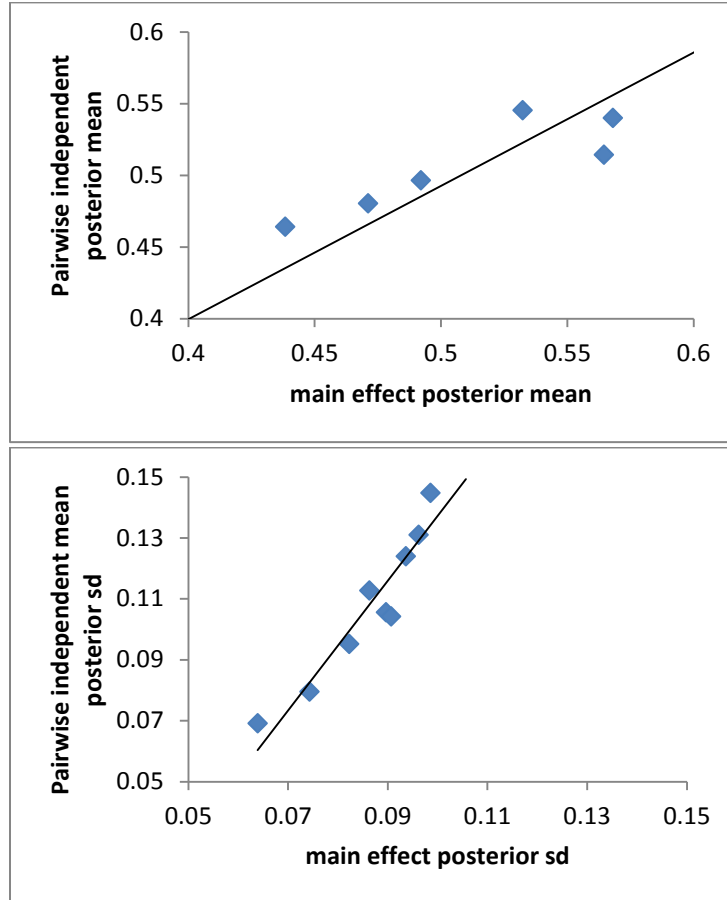
Figure 3: Posterior mean and posterior standard deviation of GOS-E from subjects across treatment arms for one simulated medium effect case for pairwise independent and main effects models.

3.1. RAR+L vs RAR vs Fixed

We present the adaptive design in terms of RAR (no longitudinal) and RAR+L (adding longitudinal modeling) over the fixed design (no early stopping, no longitudinal modeling, and no RAR) to illustrate the incremental treatment effect (Table 6). Note that the two alternative adaptive designs were recalibrated (i.e. changing what the decision rule is for success and/or

futility) so that all designs have the same Type I error of 20% and futility rate is 34% in the case of 'null hypothesis' (see Table 5). This allows for a fair comparison across designs. First, both RAR and RAR+L have a 6% and 9% power increases over the fixed design, respectively. The relative reduction of sample size in the RAR and RAR+L are 24 and 26, and the relative reduction of trial duration in the adaptive designs of RAR and RAR+L are14 and 15 weeks, respectively.

Table 6: Simulated trial operating characteristics for different designs using large effect.

Design	Power phase II	Futility Prob.	Size (n)	Duration (wks)
1. Fixed	0.88	0.00	200	140
2. RAR	0.94	0.02	176	126
3. RAR+L	0.96	0.01	174	125

4. Discussion

This manuscript presents a phase II clinical trial design that applies both pairwise independent and main effects models to explore whether one treatment for severe TBI is good enough to carry out a further phase III trial as well as to predict the probability of future success. This approach can identify the optimal treatment factor combinations in terms of pressure, duration, and addition of NBH to HBO2 for severe TBI. The pairwise independent model "drives" the RAR with longitudinal modeling and main effects model uses the data in such a trial for the phase III study prediction.

The main effects model is applied to the specific HBO2 treatment provided by the trial's goal of choosing an arm and predicting this arm's phase III success probability. Moreover, the utilization of this approach will be identical in any clinical trial in which investigators are

28

interested in adding efficiency of the prediction through the suppression of the interaction terms in the main effects. Also, this type of modeling is not limited to a binary endpoint, for example, external reviewers of our full clinical trial research plan wanted biological ties to the trial rather than solely based on clinical endpoint provided by the GOS-E. Therefore as a secondary outcome, it was proposed to measure intracranial pressure (ICP) and model this as a continuous biological response. A main effects model was also used here but rather than a binary response a normal distribution response was used. We used the expected sample sizes for power calculations for the secondary endpoints.

The investigation in this manuscript is distinct from the dose finding models where investigators might be interested in similar optimization but of a single continuous factor, and those types of trials regarding dose finding have been investigated in RAR with longitudinal modeling and have provided nice innovations to the clinical trials toolbox. However, investigators may commonly face the issue of multiple factor selection, the approach explored in this manuscript would be a good reference for them to identify the novel treatment.

The model that we propose assumes that there is no interaction between the multiple factors in the step that predicts phase III success. We use this assumption because the sample size of the phase II trial is relatively small and that the GOS-E endpoint is a noisy binary variable (e.g. relative to a continuous endpoint). Therefore it would benefit to try to find efficiencies through the statistical model. This main effects model "collapses" these factors together so that each level of pressure is providing information in three cells of the model. Of note, we do not investigate each experimental arm of treatment factors that the 3x2x2 factorial model could

29

provide, which with control arm would be 13 arms. Although it would be more efficient to keep all of the factor combinations in the design from a statistical viewpoint, the trade-off of clinical trials between risk-benefit while providing a wide enough net to grab potentially important novel treatments must be balanced. Thus, three factor combinations with too high oxygen dose were eliminated based on medical considerations. A second note is that while the phase III calculations suppress the interaction, the phase II portion uses the pairwise independent means model which actually allows for interactions of the multiple factors.

References

1. Finkelstein E, Corso P, Miller T. (2006). *The incidence and economic burden of Injuries in the United States*. New York: Oxford University Press.

2. Rockswold SB, Rockswold GL, Zaun DA, Liu J (2013), "A prospective, randomized phase II clinical trial to evaluate the effect of combined hyperbaric and normobaric hyperoxia on cerebral metabolism, intracranial pressure, oxygen toxicity, and clinical outcome in severe traumatic brain injury," *J Neurosurg,* 118(6), 1317-1328.

3. Lin K, Niu K, Tsai K, Kuo J, Wang L, Wang L, Chio C, Chang C (2012), "Attenuating inflammation but stimulating both angiogenesis and neurogenesis using hyperbaric oxygen in rats with traumatic brain injury," *J Trauma*, 72(3), 650-659.

4. Berry, SM, Spinelli, W, Littman, GS, Liang, JZ, Fardipour, P, Berry, DA, Lewis RL, and Krams, M (2010),"A Bayesian dose-finding trial with adaptive dose expansion to flexibly assess efficacy and safety of an investigational drug," *Clinical Trials*, 7, 121-135.

5. Cai, C, Liuc, S & Yuanc, Y (2014), "A Bayesian design for phase II clinical trials with

delayed responses based on multiple imputation," *Statistics in Medicine*, 33(23), 4017-4028.

6. Kim MO, Liu C, Hu F, Lee JJ (2014), "Outcome-adaptive randomization for a delayed outcome with a short-term predictor: imputation-based designs," *Statistics in Medicine*, 33(23), 4029-42.

7. Wick, J, Berry, SM, Yeh, H, Choi, W, Pacheco, CM, Daley, C, Gajewski, BJ (in press), "A Novel Evaluation of Optimality for Randomized Controlled Trials," *Journal of Biopharmaceutical Statistics*.

8. Lee, J. J., Chu, C.T. (2012), "Bayesian clinical trials in action," *Statistics in Medicine*, 31 (25), 2955-2972.

9. Connor JT, Elm JJ, Broglio KR, ADAPT-IT Investigators (2013). "Bayesian Adaptive Trials offer Advantages in Comparative Effectiveness Trials: An Example in Status Epilepticus," *Journal of Clinical Epidemiology*, 66 (8S), S130-137.

10. Berry, SM, Carlin, BP, Lee, JJ, & Muller, P (2011), *Bayesian Adaptive Methods for Clinical Trials*, CRC Press, New York.

11. Thall PF, Millikan RE, Mueller P, Lee S-J. (2003), "Dose-finding with two agents in phase I oncology trials," *Biometrics,* 59, 487–496.

12. Wang K, Ivanova A (2005), "Two-dimensional dose-finding in discrete dose space," *Biometrics*, 61, 217–222

13. Yuan Y, Yin G (2008), "Sequential continual reassessment method for two-dimensional dose-finding," *Statistics in Medicine,* 27, 5664–5678.

14. Jin, IH, Huo, L, Yin, G, Yuan, Y (2015), *"*Phase I trial design for drug combinations with Bayesian model averaging," *Pharmaceut. Statist*, 14 108–119.

15. Meurer WJ, Lewis RJ, and Berry DA, (2012), "Adaptive clinical trials: a partial remedy for the therapeutic misconception?," *Journal of the American Medical Association,* 307 (22), 2377-2378.

16. Berry, S & Sanil, A (2010), "FACTS^{TM} Dose finding: single endpoint engine specification," Tessela, Newton, MA.

17. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

18. Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**:325-337.