HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2019 June 12.

Published in final edited form as:

Biometrics. 2018 September ; 74(3): 1082–1094. doi:10.1111/biom.12841.

A Multi-source Adaptive Platform Design for Testing Sequential Combinatorial Therapeutic Strategies

Alexander M. Kaizer[1], **Brian P. Hobbs**[2], and **Joseph S. Koopmeiners**[1,*]

[1]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, U.S.A

[2]The University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.

Summary:

Traditional paradigms for clinical translation are challenged in settings where multiple contemporaneous therapeutic strategies have been identified as potentially beneficial. Platform trials have emerged as an approach for sequentially comparing multiple trials using a single protocol. The Ebola virus disease outbreak in West Africa represents one recent example which utilized a platform design. Specifically, the PREVAIL II master protocol sequentially tested new combinations of therapies against the concurrent, optimal standard of care (oSOC) strategy. Once a treatment demonstrated sufficient evidence of benefit, the treatment was added to the oSOC for all future comparisons (denoted as segments throughout the manuscript). In the interest of avoiding bias stemming from population drift, PREVAIL II considered only within-segment comparisons between the oSOC and novel treatments and failed to leverage data from oSOC patients in prior segments. This article describes adaptive design methodology aimed at boosting statistical power through Bayesian modeling and adaptive randomization. Specifically, the design uses multi-source exchangeability models to combine data from multiple segments and adaptive randomization to achieve information balance within a segment. When compared to the PREVAIL II design, we demonstrate that our proposed adaptive platform design improves power by as much as 51% with limited type-I error inflation. Further, the adaptive platform effectuates more balance with respect to the distribution of acquired information among study arms, with more patients randomized to experimental regimens.

Keywords

Adaptive randomization; Ebola virus disease; Emerging infectious diseases; Multi-source smoothing; Platform design

1. Introduction

The current translational paradigm for evaluating novel therapies is to test therapies one-at-a-time over the course of a series of two-arm trials comparing a single proposed therapy to

* koopm007@umn.edu.

SUPPLEMENTARY MATERIALS

Web Appendices and Tables referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library. R code for the simulation is available for download at https://github.com/kaize003/Multi-Source-Adaptive-Platform-Design.

an established standard of care therapy. Frequently there exists a “latency” period between each trial as the next study is designed and implemented, extending the time needed to complete the process of evaluating multiple therapies. In settings, such as oncology, wherein multiple diverse competing therapies are emerging in a contemporary period of time, master protocols facilitating trial consolidation have been proposed to limit redundancies. For example, platform trials enable multiple therapeutics to enter and exit the trial seamlessly in order to reduce the overall sample size, reduce the overall time needed to compare the multiple therapies, and improve efficiency when compared to conducting multiple independent two-arm trials (Renfro and Sargent, 2016; Hobbs et al., 2016).

The gains in efficiency from utilizing platform designs may be valuable in a general context, but can be critical in settings that require rapid evaluations of emerging therapies, such as the highly infectious Ebola virus disease (EVD) outbreak in West Africa beginning in March 2014. Initial case mortality estimates were as high as 74% and there was little to no evidence for clinical efficacy of potential therapies, which prompted the need for rapid evaluation of multiple candidate therapies (WHO, 2016; Schieffelin et al., 2014). With the need to test the clinical effectiveness of multiple agents and rapidly incorporate effective combinations into the standard of care strategy, the outbreak posed challenges to the traditional translational paradigm. In the presence of limited prior information for this particular strain of EVD as well as the potential for disease evolution, investigators were confronted with the urgent need to identify beneficial treatments or combinations of treatments as quickly as possible while maintaining traditional clinical trial benchmarks such as activity, safety, and efficacy.

In response to this need, the National Institutes of Health (NIH) launched PREVAIL II, a randomized clinical trial to evaluate medical countermeasures against EVD, in March of 2015 (Dodd et al., 2016; PREVAIL et al., 2016). A modified platform trial, PREVAIL II was designed to test multiple potentially beneficial therapeutics, accelerate clinical development, and maintain flexibility in the context of an emerging infectious disease epidemic. Driven by the urgent need for new treatments, the trial had two defining characteristics. First, the objective of PREVAIL II was to evaluate multiple treatments within a single, master protocol rather than multiple independent studies. Treatments would be evaluated sequentially against the optimal standard of care (oSOC) which, initially, consisted only of supportive care (intravenous fluids, hemodynamic monitoring, etc.). As the trial progressed, any treatment which demonstrated a significant improvement over oSOC would be added to the oSOC for future comparisons, creating a combinatorial treatment regime. Second, PREVAIL II utilized frequent interim monitoring to allow very early termination if a new combinatorial treatment regime exhibited sufficient statistical evidence for a decline in the mortality rate with respect to the concurrent oSOC.

While the PREVAIL II platform facilitated efficiency through trial consolidation, particular aspects of the statistical methodology could be enhanced for future outbreaks. One major limitation is that the proposed design only allowed the use of contemporaneous controls. For instance, if the initial drug (drug A) represented a significant improvement over the oSOC in the initial segment of the trial, the second segment would evaluate drug B in addition to drug A using simple randomization, i.e. oSOC + drug A vs. oSOC + drug A + drug B, but only the information for subjects from segment two would be used to evaluate the efficacy of

drug B, representing a redundancy which ignores the information on the oSOC + drug A experimental combination from segment one. Although the full potential of the PREVAIL II study design was not realized because the epidemic subsided before the first segment reached full enrollment and multiple treatments were not evaluated, it is important to consider how the design may be improved for future evaluations of combinatorial treatment regimes.

In this article, we describe design methodology that uses Bayesian modeling to optimize platform designs devised to test multiple combinatorial therapies over a sequence of stages by facilitating integration of information across segments. Numerous approaches exist to integrate supplemental data with data from a primary source, such as Bayesian hierarchical modeling strategies (Neuenschwander et al., 2010), commensurate priors (Hobbs et al., 2012), and multi-source exchangeability models (MEMs) (Kaizer et al., 2017). We will illustrate how MEMs and adaptive randomization can be incorporated into the PREVAIL II master protocol in order to effectuate balanced acquisition of information among the study arms and thereby maximize statistical power. Kaizer et al. (2017) illustrated that MEMs can be used to integrate information arising from potentially non-exchangeable populations while minimizing bias introduced from population drift that can occur across segments of a platform trial as patient registration trends may change over time with the incorporation of emerging experimental therapies. Integrating supplemental information using MEMs dynamically determines if the non-contemporaneous segments are exchangeable (in PREVAIL II e.g., if patient cohorts assigned to common treatment regimes exhibit evidence for equivalent mortality rates across segments) and thereby boost posterior effective sample size, in relation to the extent of evidence for bias which yields more precise estimates of the disease-response rate. Adaptive randomization (AR) can be used to balance allocation between comparator arms in relation to posterior effective sample size, increasing statistical power.

While there are numerous approaches to AR, we will consider an extension of the dynamic allocation procedure proposed by Hobbs et al. (2013), which targets information balance across treatment groups. MEMs are used to evaluate the evidence for exchangeability among current and supplemental segments as well as adapt the randomization ratio of the current segment to achieve the within-segment allocation that targets information balance between the current and supplemental control estimators and the experimental combinatorial estimator. Balancing information by boosting allocation to the novel experimental combinatorial arm improves statistical power to detect effective treatments. As effective novel therapies emerge in the platform, boosting allocation to the experimental combinatorial arm in the presence of evidence for inter-segment exchangeability among controls has the potential to improve outcomes for trial participants.

It is important to note that our proposed AR scheme differs fundamentally from conventional response- or outcome-AR methods. A recent article by Thall et al. (2015) evaluated the impact of outcome-AR, concluding that designs with outcome-AR have diminished power, necessitating a larger overall sample size. Moreover, outcome-AR risks imbalances in sample size in the wrong direction, assigning more patients to inferior treatments in the presence of the small to moderate effect sizes observed in practice. In

contrast, our proposed AR scheme endeavors to balance total effective information in the presence of potentially non-exchangeable supplemental cohorts, which maximizes power for comparing treatment groups.

The remainder of the manuscript proceeds as follows. First, the standard design of the PREVAIL II master protocol is introduced in Section 2, followed by our proposed design which incorporates MEMs and AR in Section 3. The scenarios considered for the simulation studies, the process for design calibration, and results for the simulation studies are presented in Section 4. We conclude with a brief discussion in Section 5.

2. Standard design of PREVAIL II master protocol

The objective of PREVAIL II was to sequentially evaluate multiple candidate therapies for the treatment of EVD. Each treatment was to be evaluated in a separate trial segment and each segment to consist of a separate randomized trial to compare the new treatment versus the current oSOC. Treatments found to offer a significant survival benefit when compared to the standard of care would be added to the standard of care for all future segments, creating a combinatorial treatment regime. Figure 1(I) graphically depicts an example with 3 segments and four different treatment combinations represented by color. In Segment 1, the oSOC arm (white) is compared to an experimental combinatorial arm of drug A + oSOC (light gray) with the proportion randomized to the experimental combinatorial arm fixed at $\tau = 0.5$ (represented by the equally sized triangles signifying equal enrollment throughout the segment). If the experimental combinatorial arm was determined to provide significant improvement over the standard of care, then the next trial segment would consist of a comparison between the updated oSOC, drug A + oSOC (light gray), and a new experimental combinatorial arm, drug B + drug A + oSOC (dark gray). However, if drug B + drug A + oSOC does not demonstrate a significant improvement, drug A + oSOC is carried forward to the next segment where it is compared to drug C + drug A + oSOC (darkest gray).

The PREVAIL II master protocol allowed for the rapid evaluation of multiple candidate treatments with a “Barely Bayesian” design (Proschan et al., 2016). Within a segment, PREVAIL II used frequent interim monitoring in the Bayesian paradigm to allow early termination when a treatment provided a substantial survival benefit over the standard of care (Dodd et al., 2016). The primary outcome for PREVAIL II was a binary indicator of 28-day mortality. Let x_A be the number of deaths, n_A be the total number of subjects randomized, and p_A be the 28-day mortality rate for hypothetical experimental combinatorial arm A. Define x_B, n_B and p_B analogously for hypothetical control arm B. Assuming independent beta($\alpha = 1, \beta = 1$) priors for p_A and p_B results in independent beta posteriors with $\alpha = 1 + x_A, \beta = 1 + n_A - x_A$ for p_A and $\alpha = 1 + x_B, \beta = 1 + n_B - x_B$ for p_B.

Formal inference on the 28-day mortality rate was based on the posterior distribution. The posterior probability that the 28-day mortality rate in arm A is less than arm B is:

$$P(p_A < p_B | x_A, x_B) = \sum_{k=x_A+1}^{n_A+1} \frac{\binom{n_A+1}{k}\binom{n_B+1}{x_B}(n_B - x_B + 1)}{\binom{n_A+n_B+2}{k+x_B}(n_A+n_B-k-x_B+2)}. \quad (1)$$

PREVAIL II was planned with a within-segment maximum sample size of 100 subjects per arm. If the maximum sample size were reached, the treatment A would be declared a significant improvement over the control if $P(p_A < p_B | x_A, x_B) \geq 0.975$.

PREVAIL II initiated interim analyses after six subjects were randomized to each arm and were completed after every 2 subjects until data were available for 40 subjects, whereafter interim monitoring was carried out after every 40 subjects until a maximum of 200 subjects were enrolled (100 per arm). The trial would stop and declare arm A a significant improvement over arm B if $P(p_A < p_B | x_A, x_B) \geq 0.999$. Simulation results demonstrated that this design had an overall within-segment type-I error rate near 0.03 and 86% power to detect significant difference assuming a relative risk of 0.5 for the new treatment (Dodd et al., 2016).

3. Methods

A limitation of the PREVAIL II design is that only data from contemporaneous controls were used and relevant data from prior segments was ignored. In fact, after the initial segment, non-contemporaneous, supplemental control data exists from at least one previous segment. While avoiding the introduction of inter-cohort bias due to population drift, the PREVAIL II design makes inefficient use of the data. Utilizing Bayesian methods that estimate partial exchangeability across segments overcomes this inefficiency, resulting in increased power (or decreased total sample size), while protecting against bias due to population drift.

3.1 General Framework of Multi-Source Adaptive Designs

To address this limitation, we propose the general conceptual design graphically represented in Figure 1(II). Segment 1 is identical to the original design proposed for PREVAIL II; patients are randomized with a fixed allocation ratio of $\tau = 0.5$ to the oSOC (white) or experimental drug A + oSOC (light gray). After the first segment, however, supplemental information from non-contemporaneous controls will have been acquired from past segments. This information can be integrated into future comparisons using a dynamic Bayesian model. Figure 1(II) depicts data acquired from prior segments by rectangles with diagonal lines placed in the segment of observation. For example, in Segment 2, supplemental data for the controls are available from Segment 1, which in our figure arises from the light gray study arm. In Segment 3, supplemental data for the controls are available from Segments 1 and 2.

Incorporating supplemental control information from previous segments can potentially result in imbalances in the total effective information between comparator arms if a fixed allocation ratio of $\tau = 0.5$ is maintained. Extending the adaptive randomization (AR) method proposed by Hobbs et al. (2013) to the setting of a sequential platform design attenuates this imbalance and maximizes power. This is achieved by allowing the allocation ratio to vary as a function of the effective supplemental sample size (ESSS). ESSS is a measure reflecting the extent of relative gain in the posterior precision obtained from a Bayesian model when compared to a model that neglects the supplemental sources. The measure is intended to characterize the effective number of samples incorporated from supplemental sources. By defining the allocation ratio as a function of ESSS, the proposed AR method aims to balance total information across groups within a segment. Within Figure 1(II) the allocation ratio is represented by the differing slopes, which are adjusted in relation to the extent of estimated exchangeable data contributed by concordant treatment regimes during segments 2 and 3.

The remainder of this subsection presents notation to explain the general framework for multi-source AR. During the initial period of a segment, 1:1 allocation between arms is used until sufficient information is acquired to facilitate estimation of inter-segment exchangeability, after which block-randomization is used to update the allocation ratios according to estimates of ESSS. Let n_{burn} represent the number of patients observed for the "burn-in" period at the start of a segment with supplemental information available, $\mathcal{B}$ represent the total number of blocks to adaptively randomize patients after the burn-in period, and t_b be the "time" of the b^{th} interim analysis at the start of a block. Additionally, define, at t_b, $n_A^*(t_b)$ and $n_B^*(t_b)$ as the effective sample size accounting for the number assigned in the current segment to the experimental combinatorial and control arms and the influence imparted by the prior, respectively, $ESSS(t_b)$ as the estimated ESSS for the control arm, and $R(t_b)$ as the number of subjects left to be randomized assuming the maximum segment sample size is achieved. Note that, in the context of PREVAIL II, there are no supplemental data for the experimental combinatorial arm because a treatment that illustrates a significant survival benefit will be incorporated into the oSOC arm for future segments and there were no previous human studies of the proposed therapies. Recall, the objective is to balance total effective information at trial completion such that, at t_b, $n_A^*(t_b) = ESSS(t_b) + n_B^*(t_b)$. Thus, allocation is needed in relation to $n_A^*(t_b) + \tau R = ESSS(t_b) + n_B^*(t_b) + (1 - \tau)R$. Therefore, under the the aim of balanced allocation, assignments to the experimental combinatorial arm for the next block of patients is formulated as

$$\tau(t_b) = \frac{1}{2}\left(\frac{ESSS(t_b) + n_B^*(t_b) - n_A^*(t_b)}{R(t_b)} + 1\right). \quad (2)$$

$\tau(t_b)$ can range between 0 and 1 depending on the extent of shrinkage to supplemental information with a value of 0 implying all patients are randomized to the control arm, a value of 0.5 implying a 1:1 allocation ratio, and a value of 1 implying all patients are randomized to the experimental combinatorial arm.

3.2 Incorporating supplemental information with Bayesian modeling using MEMs

Our proposed multi-source adaptive design as described thus far is general and can be enacted using any method for incorporating supplemental information. While many methods exist for incorporating supplemental information, the multi-source exchangeability model (MEM) framework is specifically considered herein based on recent efforts demonstrating its desirable properties for yielding shrinkage estimators in the presence of non- or partially exchangeable cohorts while avoiding highly parameterized models (Kaizer et al., 2017).

The MEM framework takes the H supplemental segments available for incorporation and maps them to $2^H = K$ multi-source exchangeability models, denoted $\boldsymbol{\Omega}_k$, which represent all possible exchangeability relationships between the current and H supplemental segments. For example, referring back to Figure 1(II), segment 3 would have four possible MEMs: no supplemental segments assumed exchangeable with segment 3 ($\boldsymbol{\Omega}_1$), only segment 1 assumed exchangeable with segment 3 ($\boldsymbol{\Omega}_2$), only segment 2 assumed exchangeable with segment 3 ($\boldsymbol{\Omega}_3$), and both segments 1 and 2 assumed exchangeable with segment 3 ($\boldsymbol{\Omega}_4$). The MEM framework produces a posterior estimate over these K models using posterior model weights, ω_k, such that $\Sigma_{k=1}^{K} \omega_k = 1$. A resultant smoothed posterior estimator synthesizing all possible exchangeability relationships is used for inference.

If the standard beta-binomial model is updated to accommodate MEMs in the control arm (arm B), a similar structure to the original PREVAIL II master protocol can be utilized which is able to incorporate supplemental data from previous segments, making more efficient use of available evidence and potentially improving the power of the trial. In the setting of PREVAIL II, we have supplementary data for the control arm, arm B, but not for the experimental combinatorial arm, arm A. Therefore, we will model arm A using the beta-binomial model and model arm B using MEMs. Introducing formal notation, the marginal posterior distribution of p_B given the observable data, D, from the current segment's controls and the observable data from H supplemental segments is derived as the weighted average of the posterior distributions for the K multi-source exchangeability models, $q(p_B|\boldsymbol{\Omega}_k, D)$:

$$q(p_B|D) = \sum_{k=I}^{K} \omega_k q(p_B|\boldsymbol{\Omega}_k, D). \quad (3)$$

The posterior model weight, ω_k, for each MEM is given by

$$\omega_k = pr(\boldsymbol{\Omega}_k|D) = \frac{p(D|\boldsymbol{\Omega}_k)\pi(\boldsymbol{\Omega}_k)}{\sum_{j=I}^{K} p(D|\boldsymbol{\Omega}_j)\pi(\boldsymbol{\Omega}_j)}, \quad (4)$$

where $p(D|\boldsymbol{\Omega}_k)$ is the integrated marginal likelihood for $\boldsymbol{\Omega}_k$ and $\pi(\boldsymbol{\Omega}_k)$ is the prior probability that $\boldsymbol{\Omega}_k$ is the true model. The formulation of posterior model weights in (4) utilizes a framework similar to Bayesian model averaging (BMA), however the MEM framework reduces the dimension of the prior weight space by enabling specification on the supplemental sources rather than models as described by Kaizer et al. (2017).

Using the notation from Section 2 for arm A and the current segment for arm B, let $x_{B,h}$ be the number of deaths observed in arm B for non-contemporaneous supplemental segment h ($h = 1, \ldots, H$), $n_{B,h}$ be the number of subjects randomized to arm B in segment h, and $p_{B,h}$, be the 28-day mortality rate for arm B in segment h. Let S_h denote an indicator function of whether or not the data observed in supplementary segment h is assumed exchangeable with data observed for the contemporaneous control (i.e., if $S_h = 1$, $p_{B,h} = p_B$). A model, $\mathbf{\Omega}_k$, is then defined by considering a set of source-specific binary indicators, ($S_1 = s_{1,k}, \ldots, S_H = s_{H,k}$), where $s_{h,k}$ indicates whether or not source h is assumed exchangeable with the primary data in $\mathbf{\Omega}_k$. Assuming independent beta(α,β) priors on p_B and $p_{B,1},\ldots p_{B,H}$, where each supplemental source could have different specified values for α and β, the integrated marginal likelihood for each MEM can be written as follows:

$$p(D|\mathbf{\Omega}_k) = \frac{B\left(x_B + \alpha + \sum_{h=1}^{H} s_{h,k} x_{B,h}, n_B + \beta - x_B + \sum_{j=1}^{H} s_{j,k}(n_j - x_{B,j})\right)}{B(\alpha, \beta)} \times \prod_{i=1}^{H} \left(\frac{B(x_i + \alpha, n_i + \beta - x_i)}{B(\alpha, \beta)}\right)^{1 - s_{i,k}}, \quad (5)$$

where $B(c, d) = \frac{\Gamma(c)\,\Gamma(d)}{\Gamma(c+d)}$ represents the beta function. The marginal likelihood results in the following MEM-specific posterior distribution used to calculate the marginal posterior distribution in (3):

$$q(p_B|D, \mathbf{\Omega}_k) = \text{Beta}\left(x_B + \alpha + \sum_{h=1}^{H} s_{h,k} x_h, n_B + \beta - x_B + \sum_{j=1}^{H} s_{j,k}(n_j - x_j)\right). \quad (6)$$

Therefore the posterior distribution of (3) for the 28-day mortality rate for the MEM estimator is a mixture of beta distributions encompassing all possible exchangeability relationships.

Since supplementary data are only available for the control arm, the marginal posterior probability that $p_A < p_B$ is a weighted average of the conditional posterior probability that $p_A < p_B$ for all possible assumptions about exchangeability:

$$P_{MEM}(p_A < p_B | x_A, x_B) = \sum_{i=1}^{K} \omega_i P(p_A < p_{B,\mathbf{\Omega}_i} | x_A, x_{B,\mathbf{\Omega}_i}). \quad (7)$$

Prior to implementing AR with MEMs, a 1:1 allocation ratio is assumed during the burn-in period. The specific calculations used for ESSS in (2) are defined as follows. For each individual MEM, the posterior effective sample size (ESS) can be generally derived as

$$ESS(\boldsymbol{\Omega}_k) = \alpha + \beta + n_B + \sum_{h=1}^{H} s_{h,k} n_{B,h}. \quad (8)$$

The posterior ESSS for the overall MEM estimate is then calculated as the weighted average of the difference from each individual MEM's ESS and the current control arm's sample size: $ESSS = \Sigma_{k=1}^{K} \omega_k [ESS(\boldsymbol{\Omega}_k) - n_B]$. Further, it should be noted that the beta(α,β) prior in the beta-binomial model confers the effective information of $\alpha + \beta$ subjects in (8). Therefore, the MEM facilitates a non-zero ESSS of $\alpha + \beta$ when assuming the prior probability of 1 on the independence model (e.g., the model which does not borrow strength across segments).

3.3 MEM prior probability specification

As with any Bayesian model, the properties of MEMs depend on the prior specification assumed for the model weights, with more flexible choices imparting robustness for posterior inference. Since supplemental sources are assumed independent in the MEM framework, the prior model weight formulation can be specified as the product of the source-specific prior inclusion probabilities: $\pi(\boldsymbol{\Omega}_k) = \pi(S_1 = s_{1,k}, \ldots, S_H = s_{H,k}) = \pi(S_1 = s_{1,k}) \times \cdots \times \pi(S_H = s_{H,k})$ (Kaizer et al., 2017). While there are numerous strategies to identify potential priors for each source, this section considers specific fully Bayesian and empirical Bayesian approaches which were found to achieve desirable operating characteristics in our simulation study.

Our proposed fully Bayesian prior, denoted by π_e, assumes equal prior weight for inclusion and exclusion for all supplementary sources: $\pi_e(S_h = 1) = \frac{1}{2}$. This prior provides impartiality to which supplemental segments are considered exchangeable with the primary segment.

In contrast to the fully Bayesian approach, an empirical Bayesian (EB) approach utilizes the data collected to inform the prior distribution by maximizing the marginal likelihood with respect to the prior weights. For the proposed MEM model for binary data discussed above, the marginal likelihood is maximized by placing a prior inclusion weight of 1 on sources assumed exchangeable, while all other supplemental sources receive a prior inclusion weight of 0. This induces posterior weights of 1 for the model which maximizes the marginal density and 0 for all other models. The proposed EB prior is denoted by π_{EB}.

Placing all of the weight on a single MEM, however, may induce less than ideal operating characteristics under circumstances where the marginal density of multiple MEMs may be close to the maximum marginal density. Therefore, a constrained EB prior is proposed, denoted π_{EB_c}, where $0 \leqslant c \leqslant 1$, such that the marginal density is maximized under the constraint that the prior source inclusion probabilities must be less than c. This results in a prior inclusion probability of c for segments assumed exchangeable in the MEM that maximizes the marginal density, with all other segments receiving a prior inclusion probability of 0, and the potential for multiple MEMs to receive positive prior support. When c =1, the standard EB formulation is achieved, and when c = 0, there is no borrowing

of supplemental information. Constraining the optimization over Ω_k with π_{EB_C} attenuates bias and avoids over smoothing in the presence of limited evidence for exchangeability. Section A of the Supplementary Materials provides an illustrative example of the posterior model weights under both proposed MEM priors with three supplemental segments.

3.4 Analytic Trial Example

To better illustrate the differences between the PREVAIL II study design and our proposed multi-source adaptive platform design using MEMs with the π_e and $\pi_{\mathrm{EB}10}$ priors we implement a two-segment trial example which assumes interim monitoring after every 40 subjects. Table 1 provides an example of a hypothetical trial which extrapolates the observed mortality rates from the first segment of PREVAIL II of 37% for the control arm (oSOC) and 22% in the experimental combinatorial arm (drug A) to a maximum sample size of 100 patients per arm. The results for Segment 1 are identical for all proposed approaches because no supplemental data were available that could be incorporated into Segment 1 in the context of the EVD outbreak. While the first segment does not cross the posterior probability threshold at the interim analyses, it does at the fifth and final analysis with a posterior probability of 0.9898 compared to the threshold used for the end of a segment of 0.975.

Results in Table 1 for Segment 2 are presented separately for each approach because supplemental data exist and the results will be different depending on our approach to incorporating supplemental control data. In Segment 2 we assume the new experimental combinatorial arm consists of drugs A+B and reduces the mortality rate to approximately 11% versus the 22% of drug A alone, the new control arm. It can be noted that with our proposed AR scheme, both MEM priors increase the potential number of responders randomized to the experimental combinatorial arm compared to PREVAIL II. In addition, incorporating exchangeable supplemental information increases the posterior probability of superiority due to a more precise estimate of the treatment effect, which will increase power.

4. Simulation Study

Simulation was used to evaluate and compare the operating characteristics of the PREVAIL II master protocol and the proposed multi-source AR approach. Data were generated assuming an underlying mortality rate, p_{osoc}, for oSOC alone, and the mortality rate for each potential drug combination was defined through a multiplicative model utilizing the relative risk (RR) of each drug and assuming no interactions. For example, the mortality rates for the various combinations of oSOC and two potential treatments are:

$$\begin{aligned}
&\text{oSOC} = p_{oSOC},\\
&\text{oSOC + Drug A} = p_{oSOC} \times \mathrm{RR}_A,\\
&\text{oSOC + Drug B} = p_{oSOC} \times \mathrm{RR}_B,\\
&\text{oSOC + Drug A + Drug B} = p_{oSOC} \times \mathrm{RR}_A \times \mathrm{RR}_B.
\end{aligned}$$

The multi-source AR approach assumes independent beta($\alpha = 1, \beta = 1$) priors for p_A and p_B and hypothesis testing at the interim and final analyses is based on the marginal posterior probability that p_A is less than p_B. As in PREVAIL II, our proposed design will stop and declare the experimental combinatorial arm to be a significant improvement over the control

if $P(p_A < p_B | x_A, x_B) > 0.999$ for any interim analysis. The posterior probability thresholds used at the final analysis will be calibrated for each prior to achieve the desired operating characteristics, as described in Section 4.1.

Operating characteristics are compared between PREVAIL II, the multi-source AR approach using MEMs with π_e and π_{EB_c}, and the naive approach of pooling all available supplemental information regardless of exchangeability, which will maximize the amount of supplemental information available, but introduces a prohibitive extent of bias in the presence of non-exchangeable supplementary segments. Rather than adopting the aggressive interim monitoring of PREVAIL II, we propose interim monitoring after the enrollment of every 40 subjects until the end of the burn-in period, where it will then follow the practical schedule of interim analyses at the start of each block, at t_b, with $n_{burn} = 60$ and $\mathscr{B} = 5$. This implies interim analyses after 40, 60, 95, 130, and 165 patients were observed, but updating the allocation ratio using (2) only occurs after 60, 95, 130, and 165 patients are observed. If the trial does not terminate early for superiority it will proceed to enroll a total of 200 patients in the current segment and conduct the final analysis after all information is collected. Additionally, bounds are placed on (2) to ensure $\tau(t_b) \in [0,1]$.

We considered the sequential testing of 5 potential therapeutics in the context of our platform design with two scenarios for the underlying oSOC mortality rate: (1) a constant mortality rate for all segments: $p = 0.40$ and (2) a decreasing mortality rate by segment, $p = (0.74, 0.61, 0.48, 0.36, 0.23)$. These decreasing values reflect observed mortality rates as the Ebola epidemic progressed from May 2014 to December 2014 in Sierra Leone (Dodd et al., 2016). The varying mortality scenario is more challenging for MEMs since the supplemental controls are not exchangeable, in which case minimal borrowing is preferred.

Further, five different therapeutic RR profiles are examined: (1) all drugs have a null effect (RR=1) and (2-5) one drug in the treatment pipeline has a moderate effect in segment 2, 3, 4, or 5 with RR=0.7. In a rapidly evolving epidemic it may very well be that no or very few of the included therapeutics demonstrate an improvement. Moreover, location in the pipeline may impact the platform's operating characteristics.

4.1 Design calibration

To provide context, the original PREVAIL II design had a type-I error rate of around 0.03 with 86% power to reject the null hypothesis assuming a RR of 0.5, a baseline 28-day mortality rate of 0.4, and a posterior probability threshold of 0.975 at the final analysis within a segment if the trial did not terminate early. While the same posterior probability threshold could be used for MEMs, performance will be optimized if the posterior probability threshold is optimized to achieve the desired operating characteristics. Furthermore, optimal characteristics may be achieved with a threshold that varies by segment because more supplemental information will be available at later segments as compared to earlier segments.

Given the two scenarios for the underlying mortality rate, two potential processes to calibrate the posterior probability thresholds are considered. First, thresholds can be calibrated to achieve a type-I error rate of approximately 0.025 within a segment for the

constant mortality scenario. The operating characteristics under the varying mortality scenario can be evaluated to determine if the inflation in the type-I error rate is within acceptable levels. Alternatively, thresholds can be calibrated while considering both scenarios in order to limit type-I error inflation in the varying mortality scenario while maintaining similar power in the constant mortality scenario to that observed under the PREVAIL II design.

To address the first case, potential thresholds are identified via simulation without interim monitoring using a gradient descent algorithm with $n_{burn} = 60$ and $\mathscr{B} = 5$ until the average segment-wise type-I error rate is between 0.024 and 0.026. These estimates are used as the initial values to further refine thresholds to achieve desired performance. In the second case, thresholds are identified in a segment-by-segment fashion. The aim of the second approach is to achieve power that is equal to or greater than the PREVAIL II design in the constant mortality scenario while attempting to minimize inflation of the type-I error rate in the varying mortality scenario. The latter approach may result in a trade-off between power and type-I error but, in the context of an emerging disease outbreak, enhanced power for limited type-I error inflation may be preferable. It can also be noted that the latter approach is dependent on the proposed underlying morality rates for the simulations, and that in our context we benefit from ex post facto knowledge of the epidemic. In practice, plausible scenarios need to be developed in conjunction with clinical experts with thresholds identified to balance the trade-off of type-I error and power between competing mortality profiles determined via the latter approach.

4.2 Results

25,000 simulated trials were completed for each scenario. Operating characteristics are presented for the original design of the PREVAIL II master protocol as well as the adaptive platform using MEMs with the fully Bayesian uniform prior ($\boldsymbol{\pi}_e$), MEMs with the constrained EB prior ($\boldsymbol{\pi}_{\mathrm{EB}_c}$), and naive pooling. Results are presented for thresholds calibrated to achieve the desired type-I error rate in the constant mortality scenario as described in Section 4.1. The value of $c = 0.10$ was selected for $\boldsymbol{\pi}_{\mathrm{RB}_c}$ based on extensive sensitivity analyses (not presented) that considered values of c from 0.05 to 0.50.

The operating characteristics presented for each scenario include the probability of attaining a positive test within a segment based on the Bayesian posterior probability thresholds (i.e., the probability of rejecting), the mean (sd) total number of subjects (N) treated throughout all segments as a measure of early termination, the mean (sd) proportion randomized to the experimental combinatorial arm in segments 2-5 as a measure of AR performance, and the mean (sd) proportion who survived either across segments 2-5 in the null case or in the specific non-null segment in scenarios with an efficacious therapy. When considering a drug under the null case it is ideal to rarely attain a positive test within a segment (i.e., have a probability of rejecting near 0), whereas it is desirable to attain a positive test within segments with efficacious treatments (i.e., have a probability of rejecting near 1). Further, the proportion surviving in the null scenario is expected to be identical in PREVAIL II and the proposed AR design, but an improvement in survivorship is expected in non-null

segments when the AR design effectuates more allocation to the experimental combinatorial arm.

Table 2 presents simulation results for the constant mortality scenario. The average type-I error rate across segments for MEMs is similar to or less than the average type-I error for the PREVAIL II master protocol. However, the power to detect an effective drug is higher in every non-null segment for $\pi_{\text{EB}_{10}}$ and π_e compared to PREVAIL II, with increases in power ranging from 9% to 27% and 29% to 69% for $\pi_{\text{EB}_{10}}$ and π_e, respectively. Naive pooling, which represents the best-case upper bound on performance in the presence of a constant mortality rate, results in similar or reduced type-I error rates compared to those observed in PREVAIL II with increases in power of 34% to 76% across all segments.

Operating characteristics are summarized visually in the left panel of Figure 2, where open triangles represent the results for the PREVAIL II design, closed shapes represent approaches incorporating supplemental information, and the different shades of gray identify the segment. In the presence of a constant mortality rate, all 8 MEM-based designs have increased power compared to PREVAIL II while 6 also have lower average type-I error rates.

While a moderate RR=0.7 results in minimal early termination, as observed by the average sample size estimates near 1000 for each overall trial, AR balances the information available for evaluating the control and experimental combinatorial arms, with the positive byproduct of more patients receiving a potentially beneficial treatment in the MEM-based designs than the PREVAIL design. Across segments, we observed an absolute maximum increases of 15.5% and 29.7% in the proportions assigned to the experimental combinatorial arm for $\pi_{\text{EB}_{10}}$ and π_e, respectively. This also corresponds to increases in the proportion surviving within non-null segments for the MEM-based designs compared to the PREVAIL II design with improvements observed for both $\pi_{\text{EB}_{10}}$ and π_e. Figure 3(a) presents the proportion randomized to the experimental combinatorial group and Figure 3(b) presents the proportion surviving by segment for all designs in the constant mortality scenario. π_e and naive pooling randomize a similar number to the control group, while $\pi_{\text{EB}_{10}}$ is more conservative. The MEM-based designs and naive pooling show clear increases in the median proportion surviving in each non-null scenario compared to the PREVAIL II design.

Table 3 presents simulation results for the varying mortality scenario using the thresholds calibrated for the constant mortality scenario. The PREVAIL II design maintains a type-I error rate between approximately 0.025 to 0.03 since no supplemental information is incorporated across segments, but the power steadily decreases with each segment as the underlying mortality rate continues to drop and the absolute difference in the mortality rate due to an effective treatment shrinks. Figure 2 clearly identifies that the MEM-based design with π_e (filled-in triangles) and naive pooling (filled-in diamonds) result in drastic inflation to the average type-I error rates, which negate any benefit from increased power relative to the PREVAIL II design. However, the more conservative $\pi_{\text{EB}_{10}}$ prior (filled-in squares) demonstrates a more acceptable trade-off, where the largest inflation of the type-I error rate in segment 5 increases from 0.026 for the PREVAIL II design (open triangles) to 0.064 for $\pi_{\text{EB}_{10}}$, while power increases from 2% to 51% across non-null segments. Figures 3 (c) and

(d) demonstrate similar operating characteristics to the constant mortality scenario, where the MEM-based designs and naive pooling assign higher proportions of subjects to the experimental combinatorial arm with increases in the proportion surviving the non-null segments as compared to the PREVAIL II design.

Section C of the Supplementary Materials provides additional simulation scenarios where thresholds are calibrated considering both mortality scenarios (C.1), effective therapies with RR=0.5 (C.2), varying parameters AR and MEM prior specification (C.3), and two effective therapies (C.4). Additionally, Section C.5 provides results when futility boundaries based on the posterior probability are incorporated to the trial designs.

5. Discussion

Trial consolidation with platform designs devised to sequentially evaluate multiple candidate therapies in a single trial may offer critical advantages in settings of biomedicine for which individual trials may be infeasible. With the onset of rapidly developing disease outbreaks, for example, clinical investigators are confronted with an urgent need to identify effective therapeutics as quickly as possible. During the EVD outbreak from 2014–16, the PREVAIL II study was designed to test multiple therapies within one overarching trial while incorporating aggressive interim monitoring to identify effective treatments as early as possible. The study design did not, however, utilize data observed within previous segments to improve the efficiency of the analyses, and thus is perhaps suboptimal given the availability of adaptive design features and Bayesian hierarchical modeling techniques. To address this shortcoming we proposed incorporating MEMs with AR to estimate the extent of exchangeability across segments with identical treatment regimes and balance allocation in relation to ESSS. We note, however, that this design modification could be applied with any approach to incorporating supplemental information. The methods presented herein represent a useful tool for designing platform trials not only to address future disease outbreaks, but any context where multiple contemporaneous therapeutic strategies exist.

There are many approaches to AR which could have been incorporated to our proposed design. For example, Berry and Eick (1995) compare four different response adaptive approaches to a standard equal randomization scheme and identify improvements using AR in many scenarios. More recently, Thall and Wathen (2007) explored the use of a positive constraint on Bayesian AR in order to limit the extent of adjusting the allocation ratio. Methods also exist which use biomarker information to adaptively randomize individuals during a study to more advantageous trial arms based on an individual’s biomarker profile (Zhou et al., 2008) or to adjust patient allocation in trials with binary outcomes to address covariate imbalances such that more patients can access the superior treatments identified in the study (Ning and Huang, 2010). Other methods use predictive probability and Bayesian AR to treat more patients with the more effective treatment while enabling early termination if superiority or equivalence can be demonstrated before trial completion (Yin et al., 2012).

Outcome-AR techniques, however, which can lead to sample size imbalances or poor operating characteristics are controversial (Thall et al., 2015). Our proposed design utilizes AR to target information balance, resulting in additional allocation to experimental therapies

in the presence of exchangeable information from supplemental controls. In the context of the EVD outbreak, where concerns were raised about the appropriateness of randomized trials due to ethical or practical concerns (Adebamowo et al., 2014; Ippolito et al., 2016), the modified multi-source adaptive platform design offers perhaps an ideal trade-off: controlling for cohort bias with the potential to randomize more study participants to emerging therapies.

It can be noted that 1:1 allocation is preserved in the absence of evidence of exchangeability. Further, while AR has been shown to result in potentially poor operating characteristics, MEMs with AR under $\pi_{\text{EB}_{10}}$ maintained type-I error in the constant mortality scenario with reasonable inflation in the varying mortality scenario, while increasing the power to detect an effective drug in both scenarios. This approach to AR addresses the concerns of sample size imbalance and potentially poor operating characteristics raised by Thall et al. (2015).

Even though the simulations and designs presented in this manuscript are unique to the EVD outbreak, they can be generalized to other settings of clinical study, such as a screening platform for intermediate-phased drug trials as well as sequential experimental designs of biomarker assays. In addition, a number of other parameters can be adjusted (e.g., priors on the MEM weights, the components of the adaptive randomization, and the frequency of interim monitoring) to achieve the desired operating characteristics in other settings where incorporating supplemental control information is desired.

Given the number of parameters to adjust, it may be challenging to identify the most appropriate choices given the many unknowns of a rapidly developing outbreak. However, this can be moderated by assuming conservative priors on the MEM weights, such as $c = 0.1$ for π_{EB_c}, which ensures that posterior model weight is given to the MEM assuming no exchangeable segments while incorporating information in a manner that results in an improvement to the power, survivorship, and proportion randomized to the experimental combinatorial arm. The value of c reflects a boundary imposed on the probability of exchangeability and should be evaluated in the context of each trial. Values between 0 and 1 should be explored, but in our experience c should be set at a low value to moderate the influence of π_{EB} which ultimately assigns all posterior weight to a single MEM. Further, while calibration may be challenging, the π_{EB_c} prior is based on a single hyperparameter, a benefit compared to Bayesian non-parametric density estimation with finite- or infinite-mixtures or the traditional BMA framework which requires sets of priors for all 2^H models. In addition to the MEM priors discussed in the manuscript, it is possible to set priors in-between the segments with clinician judgment to reflect the accumulated scientific knowledge regarding the relation of each supplemental segment to the current segment. We caution, however, that even the best scientific evidence cannot predict if population drift will impact future segments.

Another potential limitation is that simulation results presented in this manuscript only consider the scenario with one effective treatment across all segments. However, in the context of a rapidly emerging infectious disease, it is unlikely that an effective treatment will be present in all segments due to limited prior evaluation of the treatments. Results for cases with two effective treatments are presented in Section C.4 of the Supplementary Materials

and have similarly encouraging operating characteristics as presented in Section 4.2. Further, the results presented in this manuscript do not incorporate futility monitoring, which could limit time spent evaluating ineffective treatments. Simulation results with futility monitoring are presented in Section C.5 of the Supplementary Materials and have similar results as Section 4.2 but with slight decreases in both the average type-I error rate and power.

To this point, we have highlighted the statistical benefits of our proposed design, but a number of practical issues must be considered when implementing this design. As with any adaptive design, our proposed AR procedure is dependent on rapid endpoint acquisition. In some settings, however, patients may accrue before endpoints for previous subjects are available. In this case, methods developed for Bayesian adaptive phase I clinical trials could be used to implement AR in the presence of fast accrual (Cheung and Chappell, 2000; Koopmeiners and Modiano, 2014). In addition, real-time updates of the randomization ratio could be challenging in an emerging infectious disease epidemic. We propose to update the randomization at a fixed number of pre-planned intervals, which alleviates this concern, but the number of updates will largely be driven by practical concerns. Finally, PREVAIL II stratified randomization by country and baseline PCR cycle-threshold (PREVAIL et al., 2016). We did not consider stratified randomization to simplify the presentation of our methodology, but the relative merit of balancing information overall versus within-strata would have to be considered when implementing our AR procedure with stratified randomization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was partially funded by NIH grants P30-CA016672 and P30-CA077598.

References

Adebamowo C, Bah-Sow O, Binka F, Bruzzone R, et al. (2014). Randomised controlled trials for ebola: practical and ethical issues. Lancet 384, 1423. [PubMed: 25390318]

Berry DA and Eick SG (1995). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. Statistics in Medicine 14, 231–246. [PubMed: 7724909]

Cheung YK and Chappell R (2000). Sequential designs for phase i clinical trials with late-onset toxicities. Biometrics 56, 1177–1182. [PubMed: 11129476]

Dodd LE, Proschan MA, Neuhaus J, Koopmeiners JS, Neaton J, et al. (2016). Design of a randomized controlled trial for ebola virus disease medical countermeasures: PREVAIL II, the ebola MCM study. The Journal of Infectious Diseases 213, 1906–1913. [PubMed: 26908739]

Hobbs BP, Carlin BP, and Sargent DJ (2013). Adaptive adjustment of the randomization ratio using historical control data. Clinical Trials 10, 430–440. [PubMed: 23690095]

Hobbs BP, Chen N, and Lee JJ (2016). Controlled multi-arm platform design using predictive probability. Statistical methods in medical research.

Hobbs BP, Sargent DJ, and Carlin BP (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Analysis 7, 639–674. [PubMed: 24795786]

Ippolito G, Lanini S, Brouqui P, et al. (2016). Non-randomised ebola trials: lessons for optimal outbreak research. The Lancet Infectious Diseases 16, 407–408. [PubMed: 27036341]

Kaizer AM, Koopmeiners JS, and Hobbs BP (2017). Dynamic multi-resolution smoothing using multi-source exchangeability models. Biostatistics.

Koopmeiners JS and Modiano J (2014). A bayesian adaptive phase i–ii clinical trial for evaluating efficacy and toxicity with delayed outcomes. Clinical Trials 11, 38–48. [PubMed: 24082004]

Neuenschwander B, Capkun-Niggli G, Branson M, and Spiegelhalter DJ (2010). Summarizing historic information on controls in clinical trials. Clinical Trials 7, 5–18. [PubMed: 20156954]

Ning J and Huang X (2010). Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. Statistics in Medicine 29, 1761–1768. [PubMed: 20658546]

PREVAIL I, Team M-NPIS, et al. (2016). A randomized, controlled trial of zmapp for ebola virus infection. The New England Journal of Medicine 375, 1448–1456. [PubMed: 27732819]

Proschan MA, Dodd LE, and Price D (2016). Statistical considerations for a trial of ebola virus disease therapeutics. Clinical Trials 13, 39–48. [PubMed: 26768567]

Renfro L and Sargent D (2016). Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. Annals of Oncology 28, 34–43.

Schieffelin JS, Shaffer JG, Goba A, et al. (2014). Clinical illness and outcomes in patients with ebola in sierra leone. New England Journal of Medicine 371, 2092–2100. [PubMed: 25353969]

Thall P, Fox P, and Wathen J (2015). Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. Annals of Oncology 26, 1621–1628. [PubMed: 25979922]

Thall PF and Wathen JK (2007). Practical bayesian adaptive randomisation in clinical trials. European Journal of Cancer 43, 859–866. [PubMed: 17306975]

WHO (2016). Ebola virus disease. http://apps.who.int/mediacentre/factsheets/fs103/en/index.html Accessed: 2017-01-04.

Yin G, Chen N, and Lee JJ (2012). Phase ii trial design with bayesian adaptive randomization and predictive probability. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61, 219–235.

Zhou X, Liu S, Kim ES, Herbst RS, and Lee JJ (2008). Bayesian adaptive design for targeted therapy development in lung cancer–a step toward personalized medicine. Clinical Trials 5, 181–193. [PubMed: 18559407]

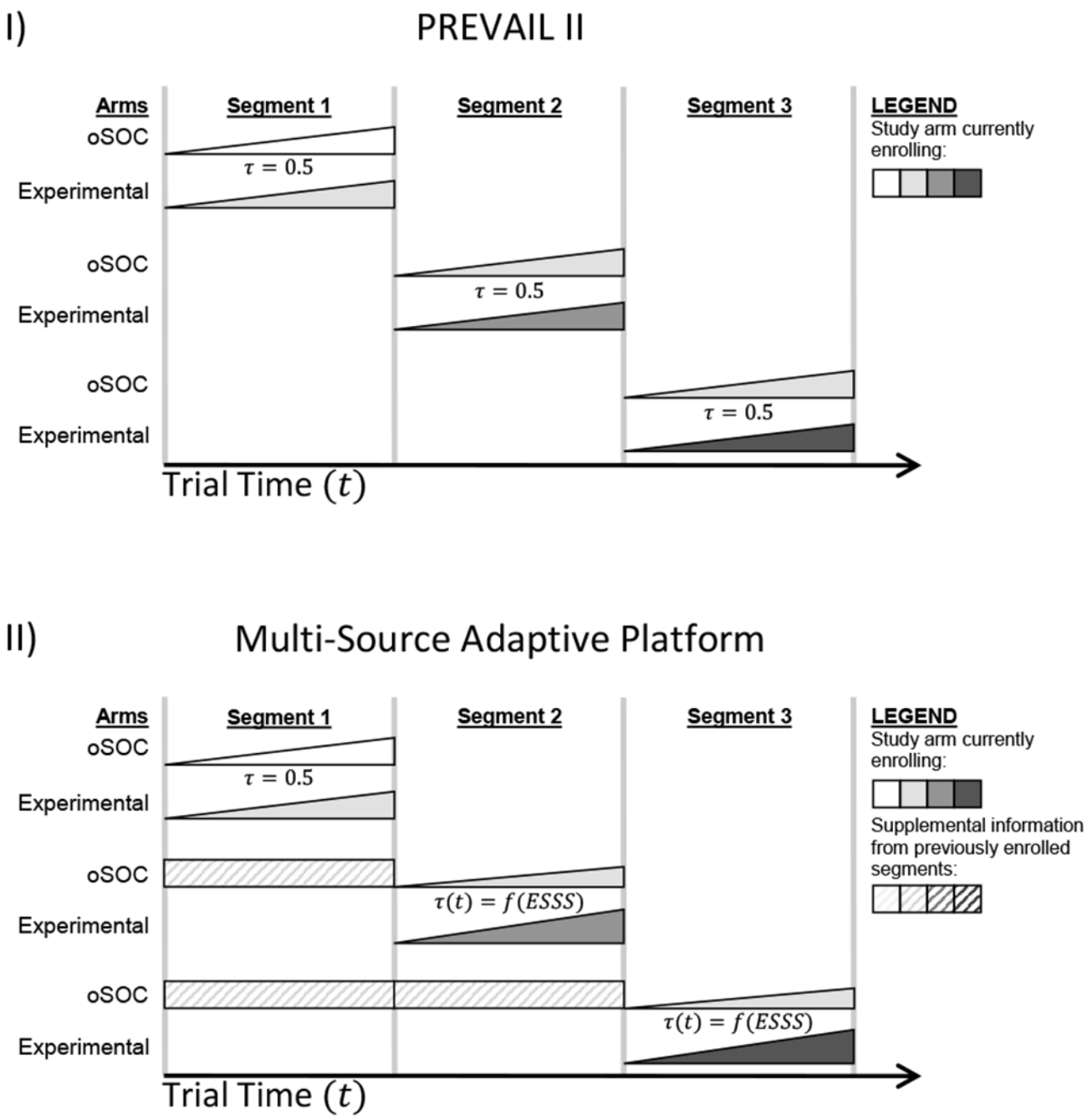

Figure 1:

Example comparing three segments of a trial with (I) PREVAIL II master protocol which only compares contemporaneously enrolled subjects with equal allocation to the study arms, $\tau = 0.5$, versus (II) framework with methods to potentially incorporate non-contemporaneous data and ability to adaptively alter the randomization ratio as a function of the effective supplemental sample size, $\pi(t) = f(ESSS)$. Equally sized triangles further indicate segments with equal allocation versus smaller oSOC triangles which indicate the potential for greater

allocation to the experimental combinatorial arm in the presence of supplemental information.

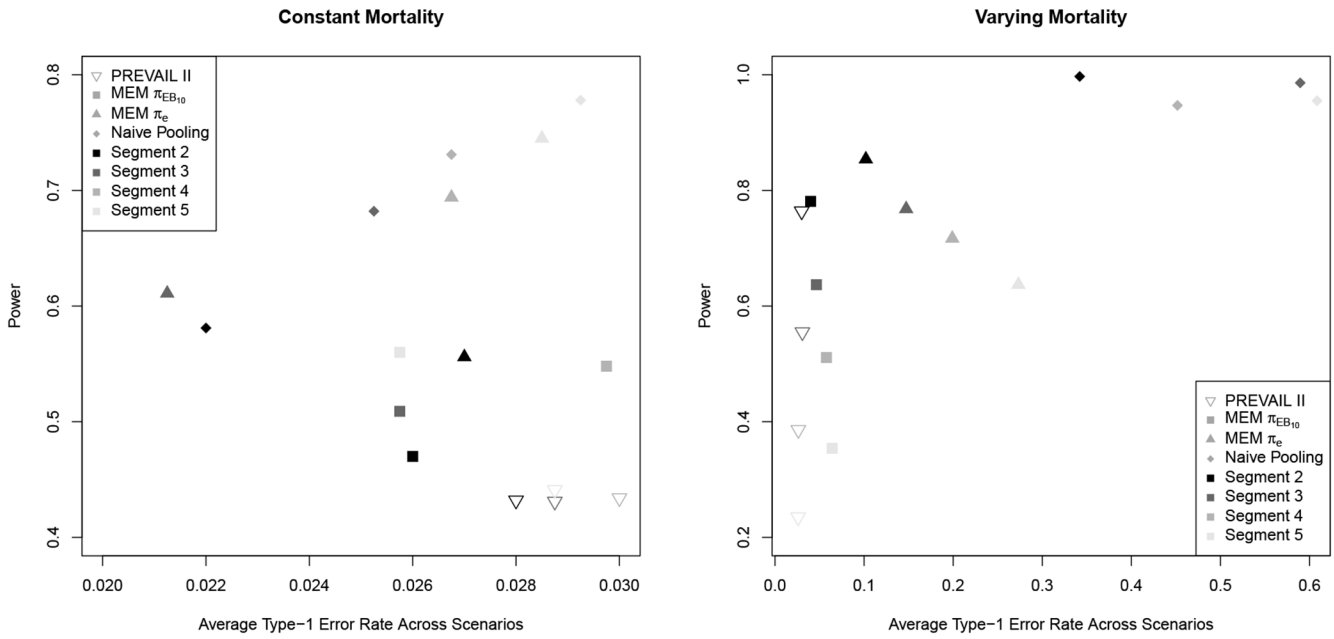

Figure 2:

Plots demonstrating power versus average type-I error rate across scenarios for segments 2-5 for PREVAIL II, MEMs with $\pi_{\mathrm{EB}_{10}}$ and π_e priors, and the naive pooling case.

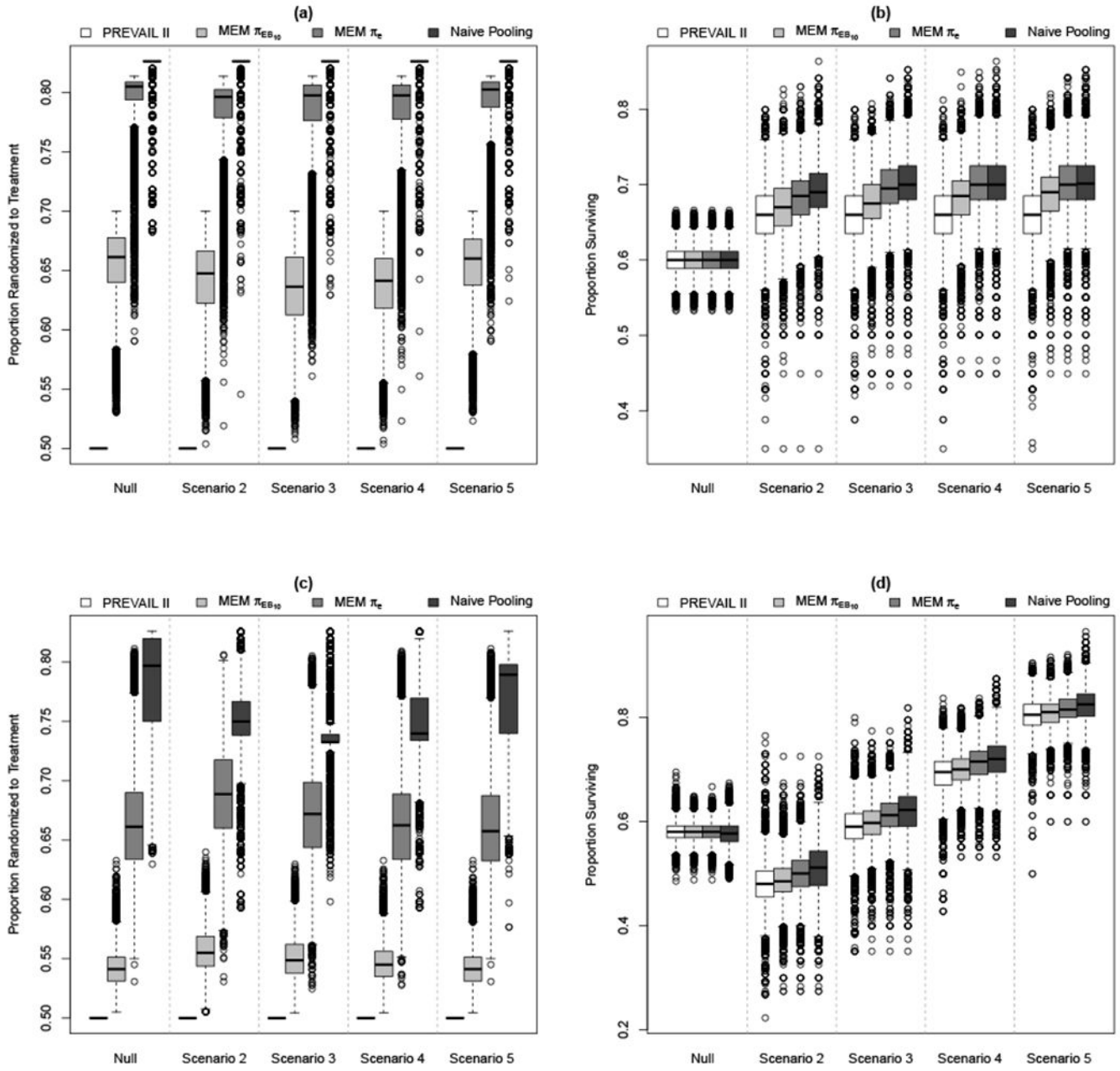

Figure 3:

Proportion assigned to experimental combinatorial arm across segments 2-5 (left) and proportion surviving across segments 2-5 (Null scenario) or within the non-null segment (Scenarios 2-5) (right) under the constant mortality scenario (top) and the varying mortality scenario (bottom) for the PREVAIL II design, MEM-based designs, and naive pooling.

Table 1:

Example analysis motivated by the PREVAIL II trial demonstrating integration of control information from Segment 1 (oSOC versus drug A) of the platform into the analysis of Segment 2 (drug A versus drugs A+B) based on the MEM approach. Mortality of 37% (oSOC group), 22% (drug A alone), and 11% (drugs A+B). Interim analyses conducted after every 40 subjects enrolled for all designs. PP stands for posterior probability that the 28-day mortality rate in the experimental combinatorial arm (drug A in Segment 1, drugs A+B in Segment 2) is less than the control arm (oSOC in Segment 1, drug A in segment 2) and τ is the proportion randomized to the experimental combinatorial arm in the subsequent segment.

Segment/Approach	Interim Analysis	n_{oSOC}	n_A	n_{A+B}	x_{oSOC}	x_A	x_{A+B}	ESSS	PP	τ
Segment 1 All Approaches	1	20	20	-	7	4	-	0	0.8471	0.5
	2	40	40	-	15	9	-	0	0.9253	0.5
	3	60	60	-	22	13	-	0	0.9634	0.5
	4	80	80	-	30	18	-	0	0.9802	0.5
	5	100	100	-	37	22	-	0	0.9898	-
Segment PREVAIL II	1	-	20	20	-	4	2	0	0.7951	0.5
	2	-	40	40	-	9	4	0	0.9298	0.5
	3	-	60	60	-	13	7	0	0.9255	0.5
	4	-	80	80	-	18	9	0	0.9699	0.5
	5	-	100	100	-	22	11	0	0.9813	-
Segment 2 MEMs with $\pi_{\text{EB}_{10}}$	1	-	20	20	-	4	2	33.2	0.8135	0.604
	2	-	36	44	-	8	5	37.8	0.9124	0.624
	3	-	51	69	-	11	7	40.7	0.9661	0.642
	4	-	65	95	-	14	10	42.5	0.9785	0.656
	5	-	79	121	-	17	13	43.8	0.9860	-
Segment 2 MEMs with π_e	1	-	20	20	-	4	2	82.3	0.8426	0.757
	2	-	30	50	-	7	6	84.1	0.9303	0.767
	3	-	39	81	-	9	9	85.5	0.9766	0.772
	4	-	48	112	-	11	12	86.5	0.9903	0.782
	5	-	57	143	-	13	16	87.3	0.9927	-

Table 2:

Operating characteristics and trial properties for the utilized platform design as well as alternative adaptive platform designs. 25,000 simulations for the constant underlying mortality case ($p = 0.4$ for all segments) with RR=0.7 for non-null segments for the PREVAIL II (P-II) master protocol; MEMs incorporating adaptive randomization with the constrained empirical Bayes, $c = .10$ prior ($\pi_{\text{EB}_{10}}$) and the fully Bayesian uniform prior (π_e); and the naive pooling (POOL) of all supplemental information incorporating adaptive randomization using posterior probability thresholds optimized for the constant mortality case. Results provided for power/type-I error for each segment, average (sd) total sample size (N) across entire trial, average (sd) proportion allocated to experimental combinatorial arm in segments 2-5, and average (sd) proportion surviving in the non-null segments (for Trt=S2-S5) or across segments 2-5 (for Trt=S0).

		Probability Reject in Segment					Mean (sd) N	Mean (sd) Prop Trt	Mean (sd) Prop Surv
	Trt	1	2	3	4	5			
PII	S0	0.032	0.028	0.029	0.031	0.029	996 (25.58)	0.5 (0)	0.600 (0.017)
	S2	0.032	0.432	0.028	0.029	0.028	988 (38.41)	0.5 (0)	0.659 (0.036)
	S3	0.032	0.028	0.431	0.029	0.030	988 (38.89)	0.5 (0)	0.659 (0.036)
	S4	0.032	0.028	0.029	0.434	0.028	988 (38.79)	0.5 (0)	0.659 (0.036)
	S5	0.032	0.028	0.029	0.031	0.441	988 (38.78)	0.5 (0)	0.659 (0.036)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.026	0.026	0.030	0.026	998 (14.97)	0.655 (0.029)	0.600 (0.017)
	S2	0.027	0.470	0.025	0.028	0.024	990 (32.03)	0.642 (0.032)	0.669 (0.035)
	S3	0.027	0.026	0.509	0.031	0.025	990 (31.49)	0.634 (0.035)	0.676 (0.035)
	S4	0.027	0.026	0.026	0.548	0.028	990 (31.09)	0.638 (0.032)	0.682 (0.035)
	S5	0.027	0.026	0.026	0.030	0.560	991 (30.96)	0.654 (0.029)	0.686 (0.036)
π_e	S0	0.027	0.027	0.026	0.033	0.037	998 (14.63)	0.797 (0.021)	0.600 (0.017)
	S2	0.027	0.556	0.017	0.020	0.026	990 (32.08)	0.785 (0.032)	0.683 (0.035)
	S3	0.027	0.027	0.611	0.021	0.024	989 (32.85)	0.782 (0.038)	0.696 (0.035)
	S4	0.027	0.027	0.026	0.694	0.027	988 (33.67)	0.784 (0.034)	0.701 (0.035)
	S5	0.027	0.027	0.026	0.033	0.745	987 (34.80)	0.794 (0.023)	0.701 (0.035)
POOL	S0	0.027	0.022	0.030	0.036	0.040	998 (15.38)	0.825 (0.009)	0.600 (0.017)
	S2	0.027	0.581	0.011	0.017	0.025	984 (37.24)	0.817 (0.024)	0.691 (0.036)
	S3	0.027	0.022	0.682	0.018	0.024	982 (38.71)	0.814 (0.030)	0.701 (0.036)
	S4	0.027	0.022	0.030	0.731	0.028	981 (38.51)	0.814 (0.028)	0.702 (0.036)
	S5	0.027	0.022	0.030	0.036	0.778	981 (38.75)	0.821 (0.013)	0.702 (0.036)

Table 3:

Operating characteristics and trial properties for the utilized platform design as well as alternative adaptive platform designs. 25,000 simulations for the varying underlying mortality case (p = (0.74, 0.61,0.48, 0.36,0.23) for segments 1-5, respectively) with RR=0.7 for non-null segments for the PREVAIL II (P-II) master protocol; MEMs incorporating adaptive randomization with the constrained empirical Bayes, c = .10 prior ($\pi_{\text{EB}_{10}}$) and the fully Bayesian uniform prior (π_e); and the naive pooling (POOL) of all supplemental information incorporating adaptive randomization using posterior probability thresholds optimized for the constant mortality case. Results provided for power/type-I error for each segment, average (sd) total sample size (N) across entire trial, average (sd) proportion allocated to experimental combinatorial arm in segments 2-5, and average (sd) proportion surviving in the non-null segments (for Trt=S2-S5) or across segments 2-5 (for Trt=S0).

		Probability Reject in Segment							
Prior	**Trt**	**1**	**2**	**3**	**4**	**5**	**Mean (sd) *N***	**Mean (sd) Prop Trt**	**Mean (sd) Prop Surv**
P-II	S0	0.028	0.030	0.032	0.027	0.025	997 (22.73)	0.5 (0)	0.580 (0.017)
	S2	0.028	0.764	0.027	0.025	0.027	972 (51.38)	0.5 (0)	0.482 (0.041)
	S3	0.028	0.030	0.555	0.026	0.026	984 (42.33)	0.5 (0)	0.591 (0.038)
	S4	0.028	0.030	0.032	0.386	0.026	990 (35.35)	0.5 (0)	0.693 (0.035)
	S5	0.028	0.030	0.032	0.027	0.235	994 (27.80)	0.5 (0)	0.804 (0.029)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.040	0.048	0.058	0.061	998 (16.20)	0.543 (0.016)	0.580 (0.017)
	S2	0.027	0.781	0.042	0.063	0.070	971 (47.98)	0.557 (0.019)	0.487 (0.038)
	S3	0.027	0.040	0.637	0.052	0.069	983 (39.13)	0.551 (0.018)	0.597 (0.036)
	S4	0.027	0.040	0.048	0.511	0.057	990 (31.19)	0.546 (0.016)	0.699 (0.032)
	S5	0.027	0.040	0.048	0.058	0.354	994 (24.99)	0.542 (0.016)	0.807 (0.027)
π_e	S0	0.027	0.102	0.165	0.213	0.262	995 (23.33)	0.665 (0.044)	0.580 (0.017)
	S2	0.027	0.854	0.094	0.236	0.326	961 (52.95)	0.689 (0.042)	0.500 (0.038)
	S3	0.027	0.102	0.768	0.134	0.307	972 (46.86)	0.673 (0.042)	0.611 (0.036)
	S4	0.027	0.102	0.165	0.717	0.198	980 (41.23)	0.665 (0.043)	0.712 (0.032)
	S5	0.027	0.102	0.165	0.213	0.637	986 (36.15)	0.663 (0.043)	0.816 (0.027)
POOL	S0	0.027	0.342	0.729	0.561	0.736	926 (54.93)	0.785 (0.038)	0.576 (0.023)
	S2	0.027	0.997	0.171	0.486	0.674	878 (48.89)	0.749 (0.034)	0.510 (0.047)
	S3	0.027	0.342	0.986	0.199	0.648	884 (39.12)	0.740 (0.029)	0.621 (0.047)
	S4	0.027	0.342	0.729	0.947	0.376	893 (44.84)	0.752 (0.036)	0.720 (0.039)
	S5	0.027	0.342	0.729	0.561	0.955	884 (56.21)	0.773 (0.039)	0.824 (0.033)