Bayesian response adaptive randomization design with a composite endpoint of mortality and morbidity

Zhongying Xu1 | Tianzhou Ma2 | Lu Tang1 | Victor B. Talisa3 | Chung-Chou H. Chang1,4

1 Department of Biostatistics, School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
2 Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, Maryland, USA
3 Department of Critical Care Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
4 Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence
Chung-Chou H. Chang, University of Pittsburgh, 200 Meyran Avenue, Suite 200, Pittsburgh, PA, USA.
Email: changj@pitt.edu

Allocating patients to treatment arms during a trial based on the observed responses accumulated up to the decision point, and sequential adaptation of this allocation, could minimize the expected number of failures or maximize total benefits to patients. In this study, we developed a Bayesian response-adaptive randomization (RAR) design targeting the endpoint of organ support-free days (OSFD) for patients admitted to the intensive care units. The OSFD is a mixture of mortality and morbidity assessed by the number of days of free of organ support within a predetermined post-randomization time-window. In the past, researchers treated OSFD as an ordinal outcome variable where the lowest category is death. We propose a novel RAR design for a composite endpoint of mortality and morbidity, for example, OSFD, by using a Bayesian mixture model with a Markov chain Monte Carlo sampling to estimate the posterior probability distribution of OSFD and determine treatment allocation ratios at each interim. Simulations were conducted to compare the performance of our proposed design under various randomization rules and different alpha spending functions. The results show that our RAR design using Bayesian inference allocated more patients to the better performing arm(s) compared to other existing adaptive rules while assuring adequate power and type I error rate control across a range of plausible clinical scenarios.

KEYWORDS
adaptive trial design, bayesian, composite endpoint, organ support free days, response adaptive randomization

1 | INTRODUCTION

To balance between individual ethics and statistical efficiency, most randomized clinical trials use fixed treatment allocation ratios in which patients are assigned among the comparative arms with fixed probability throughout the study.1 Though such fixed treatment allocation procedure remains to be the most used allocation rules in practice, more flexible and efficient allocation rules have become increasingly popular. In the response-adaptive randomization (RAR) design, treatment allocation probabilities for each arm are adjusted at different stages of the trial based on the accumulated information of treatment assignments and responses of patients already in the trial. RAR aims to achieve the objective of either maximizing patients’ total benefits by allocating more patients to receive the superior treatment or minimizing the expected number of failures for a targeted power.2 To date, numerous methods for RAR with a single type of endpoint
had been developed.2-7 Practical applications of RAR have captured much attention since the COVID-19 pandemic started.8-10

Instead of using a single type of response, many studies used composite clinical responses as the trial endpoints.11-13 In some cases, a combination of several relevant events was used to increase the combined event rate and reduce the sample size needed. However, if the components are not of equal clinical importance, that is, one event is much less severe than the other, the interpretation of such composite response may be unclear.14,15 To overcome this issue, studies have analyzed each outcome separately and separated the composite response into a primary outcome and a leading secondary outcome. Other studies treated this type of outcome as an ordinal variable and made estimation and inference accordingly. For example, we consider the outcome of interest is the organ support free days (OSFD) for critically ill patients in the intensive care units (ICUs). For each of the patient who is discharged alive from ICU, his or her OSFD is calculated up to the end of the study. If a patient dies during the ICU stay, his or her OSFD is assigned an arbitrary low number (eg, −1 day) to indicate the worst outcome. For example, if the outcome of interest was OSFD for critically ill patients in the ICUs.

In spite of the increasing development and implementation of RAR procedures in clinical trials, an RAR procedure that targets a composite endpoint of mortality and morbidity is still lacking. In this study, we propose a novel patient-benefit-oriented Bayesian response-adaptive randomization (BRAR) scheme and apply it to a multicenter, unblinded, phase II or III design with the primary endpoint of a composite endpoint of mortality and morbidity (eg, OSFD by day 28) among sepsis patients admitted to the ICUs. To handle the composite endpoint, we introduce a Bayesian mixture model to handle the composite endpoint in the RAR design. Moreover, we propose three adaptive allocation rules under the BRAR scheme. For study early stopping, unlike other RAR designs that use fixed cutoff, we consider adjusting stopping boundaries so that the overall type I error rate is under control.

In Sections 2 to 5, we illustrate the Bayesian mixture model for RAR with a composite endpoint, and demonstrate the asymptotic properties of the parameter estimators. Simulation study and the results are shown in Section 6. We summarize the work in Section 7.

2 MODEL AND PARAMETER ESTIMATE

2.1 Notation and outcome transformation

Consider an RAR trial with a total of $N$ patients and $J$ decision stages. Patients will be sequentially enrolled in the trial and randomly assigned to receive one of $K$ competing treatments. At the initial stage, roughly equal number of patients will be randomly allocated to each of the treatment arms. At each of the subsequent decision stages, treatment allocation probabilities will be adapted, and the calculations will be based on the accumulated information on treatment assignments and the responses from patients in the previous stages. Let $n_{jk}$ be the number of patients who are assigned to arm $k$ ($k = 0, 1, \ldots, K$) at stage $j$ ($j = 1, \ldots, J$), where $k = 0$ denotes the control arm. Without loss of generality, we assume an equal number of patients distributed to each stage, therefore, each stage consists of $n_j = \sum_{k=0}^{K} n_{jk} = N/J$ patients.

For each sepsis patient admitted to the ICU, we observe two types of responses: whether the patient dies during the ICU stay and how many days a patient is free of organ support if he or she is discharged alive. Denote by $\tau_{ik}$ the death indicator which equals to 1 if patient $i$ assigned to arm $k$ died during the ICU stay, and 0 otherwise. Denote by $Y_{ik}$ the number of days patient $i$ assigned to arm $k$ was not on organ support (a.k.a., the OSFD). If $\tau_{ik} = 1$, no OSFD is observed for this patient thus $Y_{ik}$ is undefined. If $\tau_{ik} = 0$, considering that OSFD is measured by Day 28 after ICU admission and OSFD greater than 28 cannot be observed, thus $Y_{ik} = \min(\text{OSFD}, 28)$. Note that $Y_{ik} \geq 1$ and can only be whole number and a larger value of $Y_{ik}$ indicates a better prognosis.

Figure 1A depicts the distribution of OSFD on the original scale $Y_{ik}$ from a hypothetical data set. The overall distribution of $Y_{ik}$ is skewed to the right and has a peak at day 28 with a large proportion of patients. $Y_{ik}$ is undefined for death events with $\tau_{ik} = 1$.

For modeling convenience, we define a random variable of composite response $D_{ik} = \{[\log(30)] - \log(30 - Y_{ik})\} \times (1 - \tau_{ik})$, where $[\log(30)]$ is the least integer greater than $\log(30)$. If a patient died by Day 28, $D_{ik} = 0$ by definition. If a patient is alive, $D_{ik} = [\log(30)] - \log(30 - Y_{ik})$ is a monotonic increasing function of $Y_{ik}$. The composite-response random variable $D_{ik} = [\log(30)] - \log(2)$ includes those patients with true OSFD of 28 and those with OSFD greater than 28. Compared to $Y_{ik}$, $D_{ik}$ (for $\tau_{ik} = 0$) is much more normally distributed with a peak at $[\log(30)] - \log(2)$ indicating the proportion of unobserved OSFD. See Figure 1B for an illustration of the impact of this transformation on the OSFD outcome.
We denote the observed data collected up to the \((j - 1)\)th stage as \(\mathcal{D}_{j-1} = \{(D_{ik}, \tau_{ik}) : j = 2, \ldots, J; i = 1, \ldots, n_{j-1,k}; k = 0, \ldots, K\}\), where \(n_{j-1,k}\) denotes the number of patients who were assigned to arm \(k\) up to stage \((j - 1)\).

### 2.2 Bayesian mixture model and prior setting

Figure 2 depicts the conceptual full hierarchical Bayesian model that can be used to model the composite endpoint of mortality and morbidity. Among patients who were discharged alive at the \(k\)th \((0 \leq k \leq K)\) arm, we consider the following probability density function of \(D_{ik}\) in our proposed two-component mixture model:

\[
\begin{align*}
 f(D_{ik} | \tau_{ik} = 0) &= \omega_k, \text{ when } D_{ik} = \log(30) - \log(2); \\
 f(D_{ik} | \tau_{ik} = 0) &= (1 - \omega_k) \cdot TN(D_{ik}; \mu_k, \sigma_k^2, [\log(30)] - \log(29), [\log(30)] - \log(2)), \text{ when } D_{ik} \neq [\log(30)] - \log(2); \\
 \tau_{ik} &\sim Ber(\lambda_k),
\end{align*}
\]

where \(TN(\mu_k, \sigma_k^2, [\log(30)] - \log(29), [\log(30)] - \log(2))\) represents the probability density function of the truncated normal distribution with mean \(\mu_k\), variance \(\sigma_k^2\), and truncated at \([\log(30)] - \log(29)\) and \([\log(30)] - \log(2)\). By definition, \(D_{ik} = 0\) when \(\tau_{ik} = 1\). \(\omega_k\) represents the proportion of unobserved OSFD peaking at \(D_{ik} = [\log(30)] - \log(2)\). Associated with \(D_{ik}\), we have a latent binary indicator variable \(Z_{ik}\), which is equal to 1 when \(D_{ik}\) is obtained from the part of unobserved OSFD and 0 otherwise. By definition from the mixture model, we have \(P(Z_{ik} = 1) = \omega_k\).

We assume the following noninformative conjugate priors for the parameters in (1):

\[
\begin{align*}
 \mu_k &\sim N(0, 10^4), & \sigma_k^2 &\sim IG(10^{-4}, 10^{-4}), \\
 \omega_k &\sim Unif(0, 1), & \lambda_k &\sim Unif(0, 1).
\end{align*}
\]
The full conditional distribution of all parameters in the model has analytic forms (see Appendix) and we use the Gibbs sampling algorithm\(^\text{16}\) to generate samples from the posterior distributions for further inference.

### 2.3 The estimand

To compare treatment effects across arms, we first estimate \(\theta_k = E(D_{ik})\), the mean transformed OSFD for arm \(k\). From (1), we have:

\[
\theta_k = E(D_{ik}) = E(D_{ik} | \tau_{ik} = 0)P(\tau_{ik} = 0) + E(D_{ik} | \tau_{ik} = 1)P(\tau_{ik} = 1)
\]

\[
= E(D_{ik} | \tau_{ik} = 0)(1 - \lambda_k) + \lambda_k
\]

\[
= \left[ \omega_k \left\{ \left( \log(30) \right) - \log(2) \right\} + (1 - \omega_k)(\mu_k - \sigma_k R_{TN,k}) \right] (1 - \lambda_k),
\]

where \( R_{TN,k} = \frac{\phi(c_{30,k}) - \phi(c_{2,k})}{\Phi(c_{30,k}) - \Phi(c_{2,k})} \) \([\cdot]\) is the ceiling function; \( c_{2,k} = \left\lceil \log(30) - \log(2) - \mu_k \right\rceil \sigma_k \); and \( \phi(\cdot) \) and \( \Phi(\cdot) \) stand for the probability density function and the cumulative distribution function of the standard normal distribution, respectively. Derivation of \( \text{Var}(D_{ik}) \), the variance of the OSFD for arm \( k \) can be found in the Appendix. We can estimate \( \theta_k \) using its posterior sample mean:

\[
\hat{\theta}_k = \frac{1}{B} \sum_{b=1}^{B} \theta_k^{(b)} / B.
\]

where \( B \) is the total number of Markov chain Monte Carlo (MCMC) iterations after the burn-in period. Based on the parameter estimates at the \( b \)th iteration and equation (3), \( \theta_k^{(b)} \) is the posterior estimate of \( \theta_k \).

We define our estimand \( \xi_k = \theta_k - \theta_0 \) as the difference in expected value of \( D_{ik} \) between the treatment arm \( k \) and the control arm \((k = 0)\). The estimand can be estimated by averaging out the posterior sample difference for all possible pairs of \( \theta \)’s:

\[
\hat{\xi}_k = \frac{1}{B^2} \sum_{b=1}^{B} \sum_{b'=1}^{B} \left( \theta_k^{(b)} - \theta_0^{(b')} \right) / B^2,
\]

where \( \theta_k^{(b)} \) and \( \theta_0^{(b')} \) are the posterior estimates of \( \theta_k \) and \( \theta_0 \) at the \( b \)th and \( (b') \)th iterations, respectively.

### 2.4 Hypothesis setting

For a study of \( K \) active treatment arms, we consider the null hypothesis of

\[
H_0 : \theta_k - \theta_0 \leq \delta, \forall k \in \{1, 2, \ldots, K\},
\]

where \( \delta \) is a clinically meaningful threshold chosen for the proposed trial of OSFD. We consider the following three alternative hypotheses:

\[
\begin{align*}
H_{a1} & : \sum_{k=1}^{K} I(\theta_k - \theta_0 > \delta) \geq 1; \\
H_{a2} & : \sum_{k=1}^{K} I(\theta_k - \theta_0 > \delta) \geq r, \text{ where } 2 \leq r \leq K - 1; \\
H_{a3} & : \sum_{k=1}^{K} I(\theta_k - \theta_0 > \delta) = K,
\end{align*}
\]
where \( I(\cdot) \) is the indicator function. The first alternative hypothesis \( H_{a1} \) aims to detect at least one best arm; the second alternative hypothesis \( H_{a2} \) aims to detect treatment effects in at least \( r \) well-performing arms; and the third hypothesis \( H_{a3} \) aims to detect treatment effect in all arms. We evaluate the posterior probability of superiority for each of the three alternative hypotheses in the simulation studies.

3 | ASYMPTOTIC PROPERTIES OF THE PARAMETER ESTIMATE

In this section, we show the asymptotic properties of the posterior mean estimates of \( \hat{\theta}_k \) and \( \hat{\xi}_k \).

**Theorem 1** (Consistency). Suppose \( D_{1k}, \ldots, D_{nk} \) are drawn i.i.d. from the correctly specified model in Equation (1) \( \{p(D_k|\Delta_k), \Delta_k \in \Theta \} \) with \( P_{\Delta_k} \) for some true parameter \( \Delta_k^* = (\lambda_k^*, \omega_k^*, \mu_k^*, \sigma_k^2) \). The true parameter \( \Delta_k^* \) is drawn from the prior \( \pi(\Delta) \) and \( \theta_k^* = g(\Delta_k^*) \) where \( g(\cdot) \) is a continuous function specified in Equation (3), then the posterior mean estimate \( \hat{\theta}_k \xrightarrow{p} \theta_k^* \).

**Proof.** The parameter \( \Delta_k = (\lambda_k, \omega_k, \mu_k, \sigma_k^2) \) in the model is of finite dimension, by Lemma 1, the Doob’s theorem (see Appendix), under the correctly specified model, the posterior mean estimate \( \hat{\Delta}_k \) is a consistent estimator, that is, \( \hat{\Delta}_k \xrightarrow{p} \Delta_k^* \). Since \( \hat{\theta}_k = g(\hat{\Delta}_k) \) is a continuous function of \( \hat{\Delta}_k \), by the continuous mapping theorem, we have \( \hat{\theta}_k \xrightarrow{p} \theta_k^* \).

**Proposition 1.** Suppose \( D_{1k}, \ldots, D_{nk} \) are drawn i.i.d. from the correctly specified model in Equation (1) \( \{p(D_k|\Delta_k), \Delta_k \in \Theta \} \) with \( P_{\Delta_k} \) for some true parameter \( \Delta_k^* = (\lambda_k^*, \omega_k^*, \mu_k^*, \sigma_k^2) \) for \( 0 \leq k \leq K \), and assuming independence between different arms. The true parameter \( \Delta_k^* \) is drawn from the prior \( \pi(\Delta) \) and \( \theta_k^* = g(\Delta_k^*) \) and \( \xi_k^* = \theta_k^* - \theta_0^* \) for \( 1 \leq k \leq K \), then the posterior mean estimate \( \hat{\xi}_k \xrightarrow{p} \theta_k^* - \theta_0^* \).

This proposition immediately follows from Theorem 1 when different arms are independent. We do not repeat the proof here.

**Theorem 2** (Asymptotic normality). For a correctly specified model in Equation (1) \( \{p(D_k|\Delta_k), \Delta_k \in \Theta \} \) with \( P_{\Delta_k} \) under certain regularity conditions and with Fisher’s information matrix \( I_{\Delta_k} \). \( \Delta_k^* \) is drawn from the prior \( \pi(\Delta) \) and \( \theta_k^* = g(\Delta_k^*) \) where \( g(\cdot) \) is a continuous function specified in Equation (3), then the posterior mean estimate follows an asymptotically normal distribution: \( \sqrt{n}(\hat{\theta}_k - \theta_k^*) \xrightarrow{d} N(0, \nabla \theta_k(\Delta_k)^T I_{\Delta_k}^{-1} \nabla \theta_k(\Delta_k)) \), where \( \nabla \theta_k(\Delta_k) \) is the gradient of \( \theta_k \) with respect to \( \Delta_k \).

We show a proof of this theorem and a derivation of the asymptotic variance in the Appendix.

**Proposition 2.** For a correctly specified model in Equation (1) \( \{p(D_k|\Delta_k), \Delta_k \in \Theta \} \) with \( P_{\Delta_k} \) under certain regularity conditions and with Fisher’s information matrix \( I_{\Delta_k} \) for \( 0 \leq k \leq K \) and assuming independence between different arms, \( \Delta_k^* \) is drawn from the prior \( \pi(\Delta) \) and \( \theta_k^* = g(\Delta_k^*) \) and \( \xi_k^* = \theta_k^* - \theta_0^* \) for \( 1 \leq k \leq K \), then the posterior mean estimate follows an asymptotically normal distribution: \( \sqrt{n}(\hat{\xi}_k - \xi_k^*) \xrightarrow{d} N\left(0, \nabla \theta_k(\Delta_k)^T I_{\Delta_k}^{-1} \nabla \theta_k(\Delta_k) + \nabla \theta_0(\Delta_0)^T I_{\Delta_0}^{-1} \nabla \theta_0(\Delta_0)\right) \), where \( \nabla \theta_k(\Delta_k) \) and \( \nabla \theta_0(\Delta_0) \) are the gradients of \( \theta_k \) and \( \theta_0 \) with respect to \( \Delta_k \) and \( \Delta_0 \), respectively.

This proposition immediately follows from Theorem 2 when different arms are independent. We also provide a derivation of the asymptotic variance in the Appendix.

4 | INTERIM ANALYSIS AND ADAPTATION RULES

The purpose of our Bayesian adaptive design is to identify the best performing treatment arm and allocate more patients to this arm. To reward the well-performing arm(s) and remove poor-performing arm(s), we consider three types of adaptation at each interim analysis stage: (a) re-estimation of the treatment allocation ratios, (b) arm suspension due to poor performance, and (b) early trial stopping due to efficacy. In the following subsections, we describe how to implement each of these adaptations in details.
4.1 Re-estimation of the treatment allocation ratios

RAR is applied here so more patients can be assigned the better treatment arm. At each interim stage \( j \), the allocation probabilities for all treatment arms \( 1 \leq k \leq K \) are updated to be proportional to the posterior probability that the arm is the best arm:

\[
p_{k,j} = P(\theta_k = \max_{1 \leq h \leq K} \theta_h | \mathcal{D}_{j-1}^+ ) = P \left( \bigcap_{h \neq k, 1 \leq h \leq K} \theta_k > \theta_h | \mathcal{D}_{j-1}^+ \right),
\]

where \( \mathcal{D}_{j-1} \) is the data collected up to the \( (j - 1) \)th stage. This posterior probability can be estimated based on the posterior samples acquired from the Bayesian model on \( \mathcal{D}_{j-1}^+ \).

With the estimated \( \hat{p}_{k,j} \), we propose the following three RAR rules depending on the proportion of patients assigned to the control arm:

**RAR Rule I:**

1. Fix the allocation proportion for the control arm to \( 1/(K + 1) \).
2. For each treatment arm \( k \) \( (1 \leq k \leq K) \), compute \( \hat{p}_{k,j} \).
3. Scale \( \hat{p}_{k,j} \) so \( \sum_{k=1}^{K} \hat{p}_{k,j} = 1 \). Allocate the remaining \( K/(K + 1) \) patients to each treatment arm proportional to the corresponding scaled probability.

**RAR Rule II:**

1. Compute the scaled \( \hat{p}_{k,j} \) as in Rule I.
2. Allocate to the best arm first, with the maximum allocation probability bounded by 0.8 to avoid extremely low numbers in the control arm.
3. For the remaining patients, allocate to the control arm with the average proportion, that is, allocate 1\( \)\( /K \) of the remaining patients to the control arm.
4. For the remaining patients, allocate to the other treatment arms with the corresponding scaled probabilities.

**RAR Rule III:**

1. Compute the scaled \( \hat{p}_{k,j} \) as in Rule I.
2. Assign patients to the control arm with the same probability as the best arm.
3. Rescale all probabilities again and allocate the remaining patients to each arm with the corresponding rescaled probabilities.

**Remark 1.** Rule I ensures that the control arm is assigned with sufficient samples; Rule II gives more rewards to the best performing arm; and Rule III well balances the best performing arm and the control arm to achieve the maximum power. Rescaling of \( \hat{p}_{k,j} \) helps avoid extreme allocation probabilities.

4.2 Arm suspension

When an arm has a very low posterior probability of being the best arm at some interim stage, we will consider suspending the arm, that is, no patients will be randomized to this arm at this stage. In our study, we consider using the threshold \( \hat{p}_{k,j} < 0.05 \) as the criteria for suspension of treatment arm \( k \) at the \( j \)th stage but other thresholds can also be assessed via sensitivity analysis. Note that the suspended arm may re-enter the study after reassessing \( \hat{p}_{k,j} \) in the next stage. The control arm is never suspended regardless of its posterior probability. For the remaining arms, we re-allocate samples following the aforementioned RAR rules.
4.3 Early stopping due to efficacy

Study will stop early for efficacy if the posterior probability of difference between the best arm and control arm greater than a clinically meaningful difference is high, that is, we evaluate the following at each interim stage $j$:

$$P(\theta_{\text{best}} - \theta_0 > \delta | \mathcal{D}_{j-1}) > c_j,$$

(7)

where $\theta_{\text{best}} = \max_{1 \leq k \leq K} \theta_k$ is the mean transformed OSFD for the best arm; $\delta$ is the clinically meaningful difference chosen in our OSFD study; and $c_j$ is the critical value used at the $j$th interim stage. The critical value is changing across stages, we propose a procedure to calculate the corresponding critical value at each stage using different alpha spending functions to control for overall type I error. We use the posterior samples acquired from the Bayesian model on $\mathcal{D}_{j-1}$ to estimate the corresponding posterior probabilities.

Denote by $n_j$ the total number of patients assigned up to the $j$th stage and $t_j = n_j / N = j/J$ the information fraction at the $j$th interim analysis. In this study, we adapt the following three alpha spending functions, where $\alpha$ is the overall type I error rate. For a one-sided test, we set $\alpha = 0.025$.

1. Pocock alpha spending function\(^{17}\):

$$\alpha(t_j) = \alpha \log\{1 + (e - 1)t_j\}.$$

2. O’Brien-Fleming (OF) alpha spending function\(^{18}\):

$$\alpha(t_j) = 2 - 2\Phi\left(\frac{z_{\alpha/2}}{\sqrt{t_j}}\right),$$

where $\Phi$ is the CDF of the standard normal distribution.

3. Power alpha spending function\(^{19}\):

$$\alpha(t_j) = (t_j)^\gamma \alpha,$$

where $\gamma$ is the power and in this study, we set $\gamma = 1$.

At each interim stage $j$, we propose the simulation-based procedure summarized in Algorithm 1 to calculate the critical value $c_j$ with respect to the alpha spending function $\alpha(t_j)$. We adopted the idea from Zhu and Yu,\(^{20}\) and proposed a revised algorithm that incorporates the characteristics of our method as follows:

4.4 Other RAR adaptation rules for comparison

We compared our proposed RAR treatment allocation rules to the following three other rules that have been used in the Bayesian RAR design with different formulas to calculate the posterior probabilities $p_{k,j}$ at each interim stage $j$ under the same Bayesian framework:

- **Fixed randomization (FR):**
  $$p_{k,j}^{\text{FR}} = 1/(K + 1) \text{ for } 0 \leq k \leq K.$$

- **Thompson sampling (TS)\(^{21}\):**
  $$p_{k,j}^{\text{TS}} = \frac{P(\theta_k = \max_{0 \leq h \leq K} \theta_{h,j} | \mathcal{D}_{j-1})^e}{\sum_{k=0}^{K} P(\theta_k = \max_{0 \leq h \leq K} \theta_{h,j} | \mathcal{D}_{j-1})^e} \text{ for } 0 \leq k \leq K,$$

where $e = \frac{n_j}{2N}$.
Algorithm 1. Algorithm to calculate the critical value $c_j$

**Data:** Simulate each of the $M_{rep}$ datasets $\mathcal{D}^m = \{(D_{ij}^m, r_{ij}^m) : i = 1, \ldots, n_j, j = 1, \ldots, J\}$, $m = 1, \ldots, M_{rep}$ for $K$ treatment arms and a control arm under $H_0 : \theta_k = \theta_0$, where $1 \leq k \leq K$ and $\theta_0$ is for the control arm. We set $M_{rep} = 10000$, $n_j = 200$, and $J = 10$.

Let $P_1 = (\tilde{P}^1, \tilde{P}^2, \ldots, \tilde{P}^{M_{rep}})'$ be a $M_{rep} \times J$ matrix,

for do $m = 1, 2, \ldots, M_{rep}$
  for do $j = 1, 2, \ldots, J$
    Calculate the posterior probability $P_j(\theta_{best} - \theta_0 > \delta | \mathcal{D}_{j-1}^m)$,
    Set $\tilde{P}_j^m = P_j(\theta_{best} - \theta_0 > \delta | \mathcal{D}_{j-1}^m)$.
  end for $j = 1, 2, \ldots, J$
end for $m = 1, 2, \ldots, M_{rep}$
for do $j = 1$
  Find $c_j = \{1 - a(t_1)\}$th quantile of the first column of the matrix $P_1$, where $t_1 = 1/J = 1/10$.
end for $j = 1$
for do $j = 2, 3, \ldots, J$
  Set $P_j$ as a matrix composed of the rows of $P_{j-1}$ such that the $(j-1)$th element of the row be smaller than or equal to $c_{j-1}$.
  Find $c_j = \{1 - \Delta a(t_j)\}$th quantile of the $j$th column of matrix $P_j$, where $\Delta a(t_j) = a(t_j) - a(t_{j-1})$.
end for $j = 2, 3, \ldots, J$

**Output:** $c_j$

• Trippa et al procedure (TP)\(^{22}\):

\[
P_{j,k}^{TP} = \frac{\pi_{k,j}}{\sum_{k=0}^{K} \pi_{k,j}}, \quad \text{where } \pi_{k,j} = \begin{cases} \frac{p(\theta_k > \theta_0 | \mathcal{D}_{j-1})^j}{\sum_{k'=0}^{K} p(\theta_{k'} > \theta_0 | \mathcal{D}_{j-1})^j}, & k = 1, \ldots, K \\ \frac{1}{K} \exp \left[ \max_{1 \leq k \leq K} r_{kj} - n_{j-1} \right], & k = 0, \end{cases}
\]

where $\gamma_j = 10(n_j/N)^{0.75}$, $\eta_j = 0.25(n_j/N)$ and $n_{kj-1}$ is the number of samples in the $k$th arm collected up to stage $j-1$.

5 | FINAL ANALYSIS AND ASSESSMENT

In the final analysis, we assessed patient benefits using the proportion of patients assigned to the best arm(s) and the posterior probability of superiority (PPS) of the trial. PPS can be served as an alternative to statistical power, the higher PPS the more powered a method is. With one treatment arm $k$, this corresponds to the difference in mean transformed OSFD between treatment arm and control arm greater than a clinically meaningful threshold $\delta$:

\[
PPS = P(\theta_k - \theta_0 > \delta | \mathcal{D}),
\]

where $\mathcal{D}$ is all data collected up to the end of the study. The PPS is estimated over $R$ Monte Carlo replications as:

\[
\hat{PPS} = \frac{1}{R} \sum_{r=1}^{R} \tilde{P}_r(\theta_k - \theta_0 > \delta | \mathcal{D}),
\]

where $R$ is the number of replications and $\tilde{P}_r(\theta_k - \theta_0 > \delta | \mathcal{D}) = \sum_{b=1}^{B} \sum_{y=1}^{B} I(\theta_k^{(y)} - \theta_0^{(b)} > \delta) / B^2$ is the posterior estimate of success probability in the $r$th replication.
In our study, we consider a total of $K$ treatment arms and one control arm. Depending on the type of the alternative hypothesis ($H_{a1}$, $H_{a2}$, or $H_{a3}$), the posterior probability of superiority can be calculated accordingly:

$$PPS(H_{a1}) = P(\theta_{\text{best}} - \theta_0 > \delta | \mathcal{D});$$

$$PPS(H_{a2}) = P \left\{ \sum_{k=1}^{K} I(\theta_k - \theta_0 > \delta) \geq r | \mathcal{D} \right\};$$

$$PPS(H_{a3}) = P(\theta_{\text{worst}} - \theta_0 > \delta | \mathcal{D}).$$

6 | SIMULATION STUDIES

6.1 | Setting

In this section, we evaluate the performance of our proposed Bayesian model and adaptation rules via simulations. We assumed a total sample size of $N = 2000$ patients in the trial. For each of the $J = 10$ interim analyses, we will perform the adaptation of the treatment allocation rule. In the simulation, we assumed a total of $K = 4$ treatment comparison arms with three active treatments and one control. For the control arm, we assumed $\mu_0 = -2.3$ and $\sigma_0 = 0.8$ for the patients and a mortality rate of $\lambda_0 = 20\%$ based on a pilot trial. Comparing to the control arm, we considered treatment arms with strong (S) effect size (Cohen’s $d = 0.8$), medium (M) effect size (Cohen’s $d = 0.5$), and no effect (N). A strong effect corresponds to mean OSFD difference of 3.5 days on the original scale, while a medium effect corresponds to mean OSFD difference of 2.5 days. We also considered different mortality rates in these arms, where $\lambda_k$ is $15\%$ for arms with strong effect, $\lambda_k$ is $18\%$ for medium, and $\lambda_k$ is $20\%$ for no effect. The standard deviations for all treatment arms were kept the same as the control arm: $\sigma_k = 0.8$, $1 \leq k \leq K$. We considered two levels of the proportion of unobserved OSFD $\omega_k \in \{0.2, 0.3\}$, $0 \leq k \leq K$ for comprehensive evaluations. We kept this proportion the same for all arms. In real data, the proportion of unobserved OSFD may vary across arms and our model is flexible to handle such variation. We simulated five scenarios with varying treatment effects as follows:

1. “MNN”: one arm has medium effect, but all other arms have no effect.
2. “SNN”: one arm has strong effect, but all other arms have no effect.
3. “SMN”: one arm has strong effect and one arm has medium effect, and the remaining third arm has no effect.
4. “SMM”: one arm has strong effect, and the other two arms have medium effects.
5. “SSM”: two arms have strong effects, and the remaining arm has medium effect.

In our simulations, we set the clinically difference $\delta = 0.67$ and $\delta = 0.65$ for $\omega_k = 0.2$ and $\omega_k = 0.3$, respectively. This corresponds to a mean OSFD difference of 1.5 days on the original scale.

We evaluated the posterior probability of superiority of the trial as well as the proportion of patients assigned to the best arm with our proposed adaptation rules. The posterior probability of superiority was assessed for all three alternative hypotheses: $H_{a1}$: at least one arm has effect; $H_{a2}$: at least two arms have effect; and $H_{a3}$: all three arms have effect. All results were based on 5000 replications of the trials. For each trial, we ran 2000 MCMC samples and removed the first 500 samples in the burn-in period, and took every 10th sample as the posterior samples for our final inference.

6.2 | Results

Figure 3 shows the proportion of patients assigned to the best performing arm(s) (with standard error) under different allocation rules in various simulation scenarios when $\omega = 0.3$. When there were two arms with strong effects, we considered both arms the best performers and calculated the sum of proportion of patients assigned to both arms. Overall, the number of arms with strong effect had a large impact on the proportion of patients assigned to the best arm, the scenarios of SSM, SNN, and MNN assigned a higher proportion of samples to the best arm(s). The proportion was higher when using our proposed adaptive allocation rules as compared to the competing rules in all scenarios. Rule II has the highest proportion among all allocation rules, around 10% higher than Rule I and 25% higher than Rule III on average. Rule II assigns patients to the best arm first to ensure sufficient samples thus has the highest proportion. Under SSM with two
Best arms, the advantage of our proposed adaptive rules over other allocation rules was becoming smaller. For the alpha spending functions, using the OBF method resulted in more patients being assigned to the best arm(s) as compared to the Pocock and power methods.

Table 1 compares the posterior probability of superiority for all scenarios under different hypotheses settings when $\omega = 0.3$. When there is only one dominating treatment arm (eg, MNN and SNN), we only evaluated $H_{a1}$. When there are more than one dominating treatment arm (eg, SMN, SMM, and SSM), we also evaluated $H_{a2}$ and $H_{a3}$. In general, scenarios with strong effect arms and more contrast between treatment arms (eg, SNN) have overall greater posterior probability of superior. In all scenarios, our proposed Rule I and Rule III are more powered than the other allocation rules for the same alternative hypothesis. Rule III achieved the greatest posterior probability of superiority in most cases with a more balanced sample assignment between the best and control arms. Though with the highest proportion assigned to the best arm(s), Rule II had the lowest posterior probability of superior among the three adaptive allocation rules we proposed, but still a similar posterior probability of superiority as the other rules, that is, fixed randomization (FR), Thompson sampling (TS), and Trippa et al procedure (TP).

7 | DISCUSSION AND CONCLUSION

In this article, we proposed a Bayesian response-adaptive randomization design for a composite endpoint combining both mortality and morbidity. We proposed a two-component mixture model for this mixed type of outcome under a Bayesian framework and used Gibbs sampling to estimate the posterior distribution of the parameters. At each interim stage, the treatment allocation probability for each arm can be estimated based on the posterior probability that the arm is the best arm given the data collected up to the current stage. We conducted extensive simulations with varying treatment effects in different arms to evaluate the adaptive allocation rules we proposed. The results showed that our proposed allocation rules are more powerful and can allocate more patients to the best arm than the other existing allocation rules. We modified the Bayesian alpha spending functions proposed by Zhu and Yu to our design and applied it to control the overall false positive error rate. Simulation results showed that more patients are assigned to the superior arm using our proposed adaptive allocation rules, with a greater posterior probability of superiority obtained at the same time.

Our Bayesian approach assumes a mixture of spike and truncated normal to model the observed endpoint based on the exploration of pilot data of the Sepsis ENdotyping in Emergency Care (SENECA) project. In any chance the distributional
| Rule | OBF | Pocock | Power | OBF | Pocock | Power | OBF | Pocock | Power | FR | TS | TP |
|------|-----|--------|-------|-----|--------|-------|-----|--------|-------|----|----|----|
| MNN  | \( H_{A1} \) | 0.806 | 0.813 | 0.810 | 0.756 | 0.762 | 0.758 | 0.829 | 0.833 | 0.826 | 0.796 | 0.809 | 0.825 |
|      | (0.225) | (0.222) | (0.225) | (0.247) | (0.248) | (0.249) | (0.212) | (0.218) | (0.219) | (0.225) | (0.225) | (0.215) |
| SNN  | \( H_{A1} \) | 0.962 | 0.961 | 0.960 | 0.936 | 0.936 | 0.936 | 0.971 | 0.972 | 0.972 | 0.932 | 0.940 | 0.949 |
|      | (0.073) | (0.075) | (0.077) | (0.114) | (0.110) | (0.114) | (0.069) | (0.055) | (0.057) | (0.097) | (0.079) | (0.078) |
| SMN  | \( H_{A1} \) | 0.957 | 0.957 | 0.956 | 0.921 | 0.927 | 0.925 | 0.966 | 0.965 | 0.965 | 0.935 | 0.937 | 0.947 |
|      | (0.085) | (0.083) | (0.088) | (0.128) | (0.126) | (0.124) | (0.072) | (0.070) | (0.073) | (0.086) | (0.080) | (0.076) |
|      | \( H_{A2} \) | 0.676 | 0.692 | 0.687 | 0.647 | 0.673 | 0.664 | 0.693 | 0.709 | 0.701 | 0.689 | 0.689 | 0.674 |
|      | (0.253) | (0.247) | (0.247) | (0.269) | (0.261) | (0.261) | (0.246) | (0.240) | (0.241) | (0.221) | (0.224) | (0.238) |
| SMM  | \( H_{A1} \) | 0.926 | 0.929 | 0.924 | 0.896 | 0.903 | 0.901 | 0.930 | 0.929 | 0.934 | 0.907 | 0.900 | 0.911 |
|      | (0.145) | (0.137) | (0.149) | (0.167) | (0.155) | (0.161) | (0.147) | (0.142) | (0.140) | (0.140) | (0.148) | (0.148) |
|      | \( H_{A2} \) | 0.743 | 0.750 | 0.749 | 0.714 | 0.724 | 0.729 | 0.755 | 0.761 | 0.764 | 0.748 | 0.731 | 0.727 |
|      | (0.222) | (0.213) | (0.219) | (0.244) | (0.236) | (0.235) | (0.218) | (0.208) | (0.210) | (0.192) | (0.205) | (0.213) |
|      | \( H_{A3} \) | 0.477 | 0.499 | 0.498 | 0.473 | 0.495 | 0.500 | 0.480 | 0.503 | 0.503 | 0.534 | 0.511 | 0.478 |
|      | (0.277) | (0.276) | (0.279) | (0.284) | (0.280) | (0.282) | (0.277) | (0.276) | (0.276) | (0.242) | (0.254) | (0.263) |
| SSM  | \( H_{A1} \) | 0.948 | 0.948 | 0.948 | 0.919 | 0.924 | 0.924 | 0.957 | 0.952 | 0.953 | 0.934 | 0.931 | 0.936 |
|      | (0.109) | (0.100) | (0.100) | (0.137) | (0.127) | (0.127) | (0.093) | (0.096) | (0.094) | (0.093) | (0.098) | (0.102) |
|      | \( H_{A2} \) | 0.822 | 0.825 | 0.823 | 0.792 | 0.803 | 0.801 | 0.837 | 0.832 | 0.831 | 0.812 | 0.806 | 0.799 |
|      | (0.201) | (0.180) | (0.185) | (0.219) | (0.207) | (0.205) | (0.187) | (0.179) | (0.181) | (0.165) | (0.171) | (0.191) |
|      | \( H_{A3} \) | 0.570 | 0.587 | 0.585 | 0.560 | 0.581 | 0.581 | 0.573 | 0.592 | 0.586 | 0.596 | 0.579 | 0.561 |
|      | (0.273) | (0.262) | (0.263) | (0.280) | (0.269) | (0.269) | (0.275) | (0.266) | (0.266) | (0.240) | (0.250) | (0.263) |

Note: \( H_{A1} \): at least one arm has treatment effect; \( H_{A2} \): at least two arms have treatment effects; \( H_{A3} \): all three arms have treatment effects. MNN: one arm has medium effect, but all other arms have no effect. SNN: one arm has strong effect, but all other arms have no effect. SMN: one arm has strong effect and one arm has medium effect, and the remaining third arm has no effect. SMM: one arm has strong effect, and the other two arms have medium effects. SSM: two arms have strong effects, and the remaining arm has medium effect. OBF: O’Brien-Fleming alpha spending function.

Abbreviations: FR, fixed randomization; TP, Trippa et al procedure; TS, Thompson sampling.

Assumption is violated, we can modify the model using nonparametric Bayesian Dirichlet mixture model or apply other popular nonparametric allocation rules such as win ratio.24

Comparing the three allocation rules proposed, Rule II has smaller posterior probability of superiority compared to all other methods, one reason for that is the actual sample size run under rule II is smaller than other two rules. Rule II assigns more patients to the better performing arm thus more likely to meet early stop criteria. Considering there is not a big gap of PPS between rule II and other two rules, we will recommend rule II in practice. In this article, we only consider the scenarios when the treatment arms are beneficial compared to the placebo. When none of the treatment arms have an effect or even harmful, the posterior probabilities of any arm being the best will accordingly drop and less patients will be assigned to any treatment arms, we will leave the investigation to future studies.

Ensuring timely data collection is critical for our study on organ support-free days in septic shock patients. Our ICU database ensures no data delays. If needed, we’ll recruit more patients to meet our sample size of 200 for each interim analysis in case of data gaps or delays within healthcare systems.

Our methods do not take patient’s baseline characteristics (eg, disease stage and race) into consideration. To further investigate the impact of patient characteristics on treatment allocation rule, we will incorporate these baseline covariates in the Bayesian model in the future development.
We developed an R package, BRACE, which is publicly available at https://github.com/joyxuuu/BRACE for implementation of our method.

AUTHOR CONTRIBUTIONS
ZX performed the analysis and wrote the manuscript. CC supervised the project and took the lead in editing the manuscript. TM, LT, and VT contributed to manuscript polishing. All authors provided critical feedback and helped to shape the research, analysis, and manuscript. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Zhongying Xu https://orcid.org/0000-0001-7883-5321
Tianzhou Ma https://orcid.org/0000-0003-3605-0811
Lu Tang https://orcid.org/0000-0001-6143-9314
Victor B. Talisa https://orcid.org/0000-0001-5246-0349
Chung-Chou H. Chang https://orcid.org/0000-0003-3665-6928

REFERENCES
1. Lin J, Lin LA, Sankoh S. A general overview of adaptive randomization design for clinical trials. JBiom Biostat. 2016;7(2):294.
2. Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal adaptive designs for binary response trials. Biometrics. 2001;57(3):909-913.
3. Wei L, Durham S. The randomized play-the-winner rule in medical trials. J Am Stat Assoc. 1978;73(364):840-843.
4. Ivanova A. A play-the-winner-type urn design with reduced variability. Metrika. 2003;58(1):1-13.
5. Villar SS, Bowden J, Wason J. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. Stat Sci. 2015;30(2):199-215.
6. Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. Ann Stat. 2004;32(1):268-301.
7. Eisele JR, Woodroofe MB. Central limit theorems for doubly adaptive biased coin designs. Ann Stat. 1995;23:234-254.
8. Angus DC. Optimizing the trade-off between learning and doing in a pandemic. JAMA. 2020;323(19):1895-1896.
9. Magaret AS, Jacob ST, Halloran ME, et al. Multigroup, adaptively randomized trials are advantageous for comparing coronavirus disease 2019 (COVID-19) interventions. Ann Intern Med. 2020;173(7):576-577.
10. Villar SS, Robertson DS, Rosenberger WF. The temptation of overgeneralizing response-adaptive randomization. Clin Infect Dis. 2021;73(3):e842.
11. Anand IS, Carson P, Galle E, et al. Cardiac resynchronization therapy reduces the risk of hospitalizations in patients with advanced heart failure: results from the comparison of medical therapy, pacing and defibrillation in heart failure (COMPANION) trial. Circulation. 2009;119(7):969-977.
12. Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. N Engl J Med. 2004;350(15):1495-1504.
13. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. N Engl J Med. 1995;333(20):1301-1308.
14. Ferreira-González I, Permanyer-Miralda G, Domingo-Salvany A, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334(7597):786.
15. Tomlinson G, Detsky AS. Composite end points in randomized trials: there is no free lunch. JAMA. 2010;303(3):267-268.
16. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984;6:721-741.
17. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977;64(2):191-199.
18. O’Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35:549-556.
19. Gordon Lan K, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-663.
20. Zhu H, Yu Q. A Bayesian sequential design using alpha spending function to control type I error. Stat Methods Med Res. 2017;26(5):2184-2196.
21. Thompson WR. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika. 1933;25(3-4):285-294.
22. Trippa L, Lee EQ, Wen PY, et al. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. J Clin Oncol. 2012;30(26):3258.
23. Annane D, Renault A, Brun-Buisson C, et al. Hydrocortisone plus fludrocortisone for adults with septic shock. N Engl J Med. 2018;378(9):809-818.
24. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176-182.
25. Doob JL, Doob JL. *Stochastic Processes*. Vol 7. New York: Wiley; 1953.
26. van der Vaart AW. *Asymptotic Statistics*. Vol 3. Cambridge, UK: Cambridge University Press; 2000.

**SUPPORTING INFORMATION**

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

**How to cite this article:** Xu Z, Ma T, Tang L, Talisa VB, Chang C-CH. Bayesian response adaptive randomization design with a composite endpoint of mortality and morbidity. *Statistics in Medicine*. 2024;43(6):1256-1270. doi: 10.1002/sim.10014

**APPENDIX A. FULL CONDITIONAL POSTERIOR DISTRIBUTION FOR GIBBS SAMPLING**

We used Gibbs sampling to update all parameters in the model, the corresponding full conditional posterior distributions for each parameter are summarized below.

\[
(\lambda_k | \mu_k, \sigma^2_k, \omega_k, z_{ik}, D_{ik}, \tau_{ik}) \sim \text{Beta}(1 + \sum_i r_{ik}, 1 + n - \sum_i r_{ik}),
\]

\[
(\omega_k | \lambda_k, \mu_k, \sigma^2_k, z_{ik}, D_{ik}, \tau_{ik}) \sim \text{Beta}(1 + \sum_i z_{ik}, 1 + n - \sum_i z_{ik}),
\]

\[
(z_{ik} | \lambda_k, \mu_k, \sigma^2_k, \omega_k, D_{ik}, \tau_{ik}) \sim \text{Ber}\left[\frac{\omega_k}{\omega_k + (1 - \omega_k) \times TN[D_{ik} | \mu_k, \sigma^2_k, [\log(30)] - \log(30), [\log(30)] - \log(2)]}\right],
\]

\[
(\mu_k | \lambda_k, \sigma^2_k, \omega_k, z_{ik}, D_{ik}, \tau_{ik}) \sim N\left(\frac{1}{10^n} + \frac{n}{\sigma^2_k}, \frac{1}{10^n} + \frac{n}{\sigma^2_k}\right),
\]

\[
(\sigma^2_k | \lambda_k, \mu_k, \omega_k, z_{ik}, D_{ik}, \tau_{ik}) \sim IG\left(10^{-4} + \frac{n}{2}, 10^{-4} + \frac{\sum_{i=1}^n (D_{ik} - \mu_k)^2}{2}\right).
\]

**APPENDIX B. DERIVATION OF VARIANCE OF TRANSFORMED OSFD**

The variance of \(D_{ik}\) can be derived as follows:

\[
\text{Var}(D_{ik}) = E[\{D_{ik} - E(D_{ik})\}^2 | \tau_{ik} = 0]P(\tau_{ik} = 0) + E[\{D_{ik} - E(D_{ik})\}^2 | \tau_{ik} = 1]P(\tau_{ik} = 1)
\]
\[= \left[\text{Var}(D_{ik}^2 | \tau_{ik} = 0) - E^2(D_{ik} | \tau_{ik} = 0)\right](1 - \lambda_k) + 0 \cdot \lambda_k
\]
\[= \omega_k \left(\left[\log(30) - \log(2)\right]^2 + (1 - \omega_k) \left(\text{Var}_{TN(\mu_k, \sigma^2_k)} + \text{Var}_{TN(\mu_k, \sigma^2_k)}\right) - \theta_k^2\right)(1 - \lambda_k)
\]
\[= \omega_k \left(\left[\log(30) - \log(2)\right]^2 + (1 - \omega_k) \left\{\left(\mu_k - \sigma_k R_{TN,k}\right)^2 + \sigma^2_k \left(1 + W_{TN,k} - R_{TN,k}^2\right)\right\}
\]
\[- \left\{\left[\log(30) - \log(2)\right]^2 + (1 - \omega_k) (\mu_k - \sigma_k R_{TN,k})^2\right\}(1 - \lambda_k)\right].
\]

where \(R_{TN,k} = \frac{\phi(c_{2,k}) - \phi(c_{30,k})}{\phi(c_{2,k}) - \phi(c_{30,k})}; W_{TN,k} = c_{2,k} \phi(c_{2,k}) - c_{30,k} \phi(c_{30,k})\); \(c_{2,k} = \frac{\left[\log(30) - \log(2) - \mu_k\right]}{\sigma_k}\); and \(c_{30,k} = \frac{\left[\log(30) - \log(29) - \mu_k\right]}{\sigma_k}\).
**APPENDIX C. PROOF OF THEOREMS**

**Lemma 1.** Doob’s consistency theorem\(^{25}\). Suppose \(X_1, \ldots, X_n\) are i.i.d. drawn from a correctly specified model \(\{p(X|\gamma), \gamma \in \mathbb{Y}\}\) with \(P_\gamma\) for some true parameter \(\gamma^*\) and \(\gamma^*\) is drawn from the prior \(\pi(\gamma)\), then the posterior mean \(\hat{\gamma} \overset{p}{\longrightarrow} \gamma^*\).

**Remark 2.** A proof of this theorem is provided in Reference 26, we do not repeat it here. Doob’s consistency theorem implies that the posterior mean of a parameter with finite dimension in a correctly specified model is consistent.

**Lemma 2.** Bernstein-von Mises theorem\(^{26}\). For a well-specified regular parametric model \(\{p(X|\Delta), \gamma \in \mathbb{Y}\}\) with \(P_\gamma\), under certain regularity conditions and with nonsingular Fisher information matrix \(I_\gamma\), suppose the prior density \(\pi(\gamma)\) is continuous in a neighborhood of \(\gamma^*\), then the corresponding posterior law satisfy:

\[
||P_n(\hat{\gamma} - \gamma^*)||_{||} \overset{N(0, \Gamma^{-1}_\gamma)}{\longrightarrow} 0.
\]

As \(n \to \infty\), \(v_n, \gamma^*\) weakly converges to \(N(0, \Gamma^{-1}_\gamma)\).

**Remark 3.** A proof of this theorem is provided in Reference 26. We do not repeat it here. The Bernstein-von Mises theorem links Bayesian inference with frequentist inference. It shows that the posterior laws converge in distribution to a Gaussian posterior law in total variation distance. This immediately implies that any location functional suitably continuous relative to the total variation norm applied to the sequence of posterior laws converges to the same location functional applied to the limiting Gaussian posterior distribution. For most choices, this means \(N(0, \Gamma^{-1}_\gamma)\). Suppose the posterior mean is defined as \(\hat{\gamma} = \int \gamma p(\gamma|X)d\gamma\). An immediate result from the Bernstein-von Mises theorem implies the asymptotic normal distribution of \(\hat{\gamma}\), that is, \(\sqrt{n}(\hat{\gamma} - \gamma^*) \sim N(0, \Gamma^{-1}_\gamma)\). Basically, the Bernstein-von Mises theorem establishes that \(\hat{\gamma}\) is on the same footing as the maximum likelihood estimate of \(\gamma\).

### C.1 Proof of Theorem 2

**Proof.** The model meets the general regularity conditions and the prior \(\pi(\Delta)\) is continuous, so the asymptotic normality results of \(\hat{\Delta}_k\) directly follow the Bernstein-von Mises theorem, that is, \(\sqrt{n}(\hat{\Delta}_k - \Delta^*) \overset{d}{\longrightarrow} N(0, \Gamma^{-1}_\Delta)\). What we show here is the derivation of the nonsingular Fisher’s information matrix thus the asymptotic variance.

We first let \(\Delta_k = (\lambda_k, \omega_k, \mu_k, \sigma^2_k)\) and write out the full likelihood of the model:

\[
\mathcal{L}(\Delta_k|D_{ik}, \tau_{ik}) = \prod_{i=1}^{n} f(D_{ik}|\tau_{ik})
= \prod_{i=1}^{n} \left\{ f(D_{ik}|\tau_{ik} = 0)P(\tau_{ik} = 0) + f(D_{ik}|\tau_{ik} = 1)P(\tau_{ik} = 1) \right\}
= \prod_{i=1}^{n} (1 - \lambda_k) \left\{ \omega_k + (1 - \omega_k)TN(D_{ik}; \mu_k, \sigma^2_k, [\log(30)] - \log(29), [\log(30)] - \log(2)) \right\}.
\]

Taking the log, we get the log-likelihood:

\[
\ell(\Delta_k|D_{ik}, \tau_{ik}) = n \log(1 - \lambda_k) + \sum_{i=1}^{n} \log \left\{ \omega_k + (1 - \omega_k)TN(D_{ik}; \mu_k, \sigma^2_k, [\log(30)] - \log(29), [\log(30)] - \log(2)) \right\},
\]

where \(TN(D_{ik}; \mu_k, \sigma^2_k, [\log(30)] - \log(29), [\log(30)] - \log(2)) = \frac{\phi(\frac{D_{ik} - \mu_k}{\sigma_k})}{\phi(\frac{\log(30) - \log(29)}{\sigma_k}) - \phi(\frac{\log(30) - \log(29) - r_k}{\sigma_k})} \) \(\Phi(\cdot)\) and \(\Phi(\cdot)\) stand for the PDF and CDF of the standard normal distribution, respectively.
We then derive the score function:

\[
S(\Delta_k) = \begin{pmatrix}
\frac{\partial r(\Delta_k)}{\partial \alpha_k} \\
\frac{\partial r(\Delta_k)}{\partial \beta_k} \\
\frac{\partial r(\Delta_k)}{\partial \gamma_k} \\
\frac{\partial r(\Delta_k)}{\partial \delta_k}
\end{pmatrix} = \left( -\frac{\alpha_k}{\alpha_k} \frac{1}{\alpha_k + (1 - \alpha_k)\psi(D_{\alpha_k})} \right),
\]

where \( \psi(.) \) is the PDF of \( TN(\mu_k, \sigma^2_k, [\log(30)] - \log(29), [\log(30)] - \log(2)) \), \( \psi'(D_{\alpha_k}) = \frac{\partial \psi(D_{\alpha_k})}{\partial \mu_k} \), \( \psi'(D_{\alpha_k}) = \frac{\partial \psi(D_{\alpha_k})}{\partial \sigma^2_k} \).

We also know for each \( i \)-th case, \( S(\Delta_k) \) = \( \begin{pmatrix}
-\frac{1}{\alpha_k} \frac{1}{\alpha_k + (1 - \alpha_k)\psi(D_{\alpha_k})} \frac{(1 - \alpha_k)\psi'(D_{\alpha_k})}{\sigma^2_k} \\
0 \\
0 \\
0
\end{pmatrix} \).

Assuming i.i.d., we can further derive Fisher’s information matrix:

\[
I(\Delta_k) = -E\left\{ \frac{\partial S(\Delta_k)}{\partial \Delta_k} \right\} = E \begin{pmatrix}
-\frac{1}{\alpha_k} & 0 & 0 & 0 \\
0 & I_{1,22} & I_{1,23} & I_{1,24} \\
0 & I_{1,23} & I_{1,33} & I_{1,34} \\
0 & I_{1,24} & I_{1,34} & I_{1,44}
\end{pmatrix},
\]

where \( \psi_i = \psi(D_{\alpha_k}), \psi_i' = \psi'(D_{\alpha_k}), \psi_i'' = \psi''(D_{\alpha_k}), \psi_i''' = \psi'''(D_{\alpha_k}), \psi_i(1 - \alpha_k) = \frac{\partial \psi(D_{\alpha_k})}{\partial \mu_k}, \psi_i(1 - \alpha_k) = \frac{\partial \psi(D_{\alpha_k})}{\partial \sigma^2_k}, \psi_i = \frac{\partial \psi(D_{\alpha_k})}{\partial \alpha_k}, \psi_i = \frac{\partial \psi(D_{\alpha_k})}{\partial \gamma_k} \).

We have completed the derivation of a non-singular information matrix \( I(\Delta_k) \).

Since \( \hat{\theta}_k = g(\Delta_k) \) is a continuous function of \( \Delta_k \), we can apply the multivariate Delta’s method and obtain the asymptotic normal distribution of \( \hat{\theta}_k \): \( \sqrt{n}(\hat{\theta}_k - \theta^*_k) \xrightarrow{d} N\left\{ 0, \nabla \theta_k(\Delta_k)^T \Gamma^{-1}_k \nabla \theta_k(\Delta_k) \right\} \).

What we need to do here is just to derive \( \nabla \theta_k(\Delta_k) \):

\[
\nabla \theta_k(\Delta_k) = \begin{pmatrix}
\frac{\partial \theta_k}{\partial \alpha_k} \\
\frac{\partial \theta_k}{\partial \beta_k} \\
\frac{\partial \theta_k}{\partial \gamma_k} \\
\frac{\partial \theta_k}{\partial \delta_k}
\end{pmatrix} = \begin{pmatrix}
-\frac{1}{\alpha_k} & 0 & 0 & 0 \\
0 & I_{1,22} & I_{1,23} & I_{1,24} \\
0 & I_{1,23} & I_{1,33} & I_{1,34} \\
0 & I_{1,24} & I_{1,34} & I_{1,44}
\end{pmatrix},
\]

where \( R_{TN,k} = \frac{\theta_k}{\frac{[\log(30)] - \log(29)}{\sigma^2_k} \frac{[\log(30)] - \log(2)}{\sigma^2_k}} - \frac{\theta_k}{\frac{[\log(30)] - \log(29)}{\sigma^2_k} \frac{[\log(30)] - \log(2)}{\sigma^2_k}}, \) \( R_{TN,k}^1 = \frac{\partial R_{TN,k}}{\partial \alpha_k}, \) \( R_{TN,k}^2 = \frac{\partial R_{TN,k}}{\partial \gamma_k} \).