Published in final edited form as:

Pharm Stat. 2018 July ; 17(4): 342–357. doi:10.1002/pst.1851.

Bayesian approach for assessing non-inferiority in a three-arm trial with Binary endpoint

Santu Ghosh¹, Ram C. Tiwari², and Samiran Ghosh^{3,4}

¹Department of Biostatistics and Epidemiology, Augusta University

²Division of Biostatistics, CDRH, FDA, USA

³Department of Family Medicine & Public Health Sciences

⁴Center of Molecular Medicine and Genetics, Wayne State University

Abstract

With the recent advancement in many therapeutic areas, quest for better and enhanced treatment options are ever increasing. While the “efficacy” metric plays the most important role in this development, emphasis on other important clinical factors such as less intensive side effects, lower toxicity, ease of delivery and other less debilitating factors may result in the selection of treatment options which may not beat current established treatment option in terms efficacy, yet prove to be desirable for subgroups of patients. The resultant clinical trial by means of which one establishes such slightly less efficacious treatment is known as non-inferiority (NI) trial. NI trials often involve an active established comparator arm, along with a placebo and an experimental treatment arm, resulting into a three-arm trial. Most of the past developments in a three-arm NI trial consider defining a pre-specified fraction of unknown effect size of reference drug, i.e., without directly specifying a fixed non-inferiority margin. However, in some recent developments more direct approach is being considered with pre-specified fixed margin, albeit in the frequentist setup. In this article we consider Bayesian implementation of such trial when primary outcome of interest is binary. Bayesian paradigm is important as it provides a path to integrate historical trials and current trial information via sequential learning. We employ several approximation-based as well as two exact fully-Bayesian methods to evaluate the feasibility of the proposed approach. Finally, a clinical trial example is reanalyzed to demonstrate the benefit of the proposed approach.

Keywords

Assay Sensitivity; Bayesian Method; Jeffreys’ prior; Markov chain Monte Carlo; Noninferiorty margin

1 Introduction

Randomized controlled trials are gold standard method to establish a new treatment regime for any disease area. In the absence of an established treatment option, the usefulness of

placebo controlled trial is unquestionable. However, when there exists an established treatment regime, called the standard treatment, withholding that in favor of placebo controlled trial is unethical and can lead to serious and irreversible adverse events. This leads to a class of trials, known as active control trial where an experimental drug is compared against one or more established treatment regimes, referred to as an active control or reference. Comparative Effectiveness Research (CER) is a good example where active control trials are more of a norm. For simplicity of exposition, we consider here one test drug versus one active control; however, multiple comparisons are very much possible (e.g., see Kwong et al. [1]). When superiority of the new test (experimental) treatment is questionable, we often test to see if the test treatment is not worse than an active control (reference) by more than a specified margin. For example, a trial might seek to determine if an inexpensive and/or easy to administer and/or less toxic test treatment maintains a substantial portion of the effect of the active control [2]. This leads to a special type of active control trial often known as non-inferiority (NI) trial. The objective of this kind of trial is to define first an acceptable loss of efficacy, often termed as non-inferiority margin (> 0) and then to verify whether the difference between the test drug and active control drug falls within this margin. As expected the choice of allowable loss of efficacy (i.e., the NI margin) is a rather critical issue. As we were finishing this article, FDA on November 7th, 2016 finalized its 2010 draft guidance [3; 4] to pharmaceutical and biotech sponsors looking for more advice on when non-inferiority studies demonstrating effectiveness of an investigational drug can provide interpretable results and how to choose the non-inferiority margin. It should be noted various other agencies also provided broad guidelines [5; 6; 7] in the past on the choice of NI margin, however a consensus solution is still matter of debate. Hence, the NI trials need to be administered with extreme caution, a detailed discussion of which can be found in the references [4; 8; 9].

In the last few decades, the literature dealing with NI trials has grown substantially albeit from frequentist paradigm. Most of these methods consider two-arm trials consisting of experimental (E) and reference (R) treatments. This makes ethical sense as no one is exposed to the placebo (P), however, such a trial has some major challenges in terms of design, analysis, and possible interpretation of results. For example, due to the absence of placebo arm the two-arm NI trials are unable to establish direct proof of efficacy of (i) the reference treatment over placebo and (ii) the experimental treatment over placebo, and require external validation which is often questionable. Hence, two-arm non-inferiority trials, without placebo, lack the support of assay sensitivity (ICHE10 [6], Chapter 1.5). Assay sensitivity refers to the ability to distinguish between effective and ineffective treatments. Hence, a NI trial should not only show comparative effectiveness but also demonstrate that the active control has similar effect size that was reported in historical placebo-controlled studies. As reported in Kieser and Friede [10] this requires strong assumptions such as a) availability of reliable historical placebo-controlled trial for reference, b) absence of publication bias, c) absence of selection bias, d) spatio-temporal constancy of trial design, and e) unaltered clinical practice since the establishment of reference over placebo. Establishing all of these is a near impossible task. Hence to alleviate some of these problems, the inclusion of a placebo arm is recommended in a NI trial by EMA [7], when it is feasible and ethically justified. The resulting three-arm trial (with E, R

and P arms) is also known as “*gold standard design*” and is preferable due to its greater confidence concerning assay sensitivity (AS) and lesser concern related to external validity.

As mentioned earlier a major point of any NI trial is the choice of NI margin, δ . For three-arm trials, Pigeot et al. [11] and Koch and Rohmel [12] proposed the choice of NI margin as the prespecified fraction of the unknown effect size of reference drug, i.e., without directly specifying a fixed non-inferiority margin. This is also known as effect retention hypothesis testing approach where non-inferiority hypothesis is tested as the ratio of differences of normal outcomes for continuous data. Kieser and Friede [10] extended the approach for binary end-point by considering difference of proportions. Though Pigeot et al. [11] mentioned clearly that such NI testing must be preceded by an assay sensitivity testing where one should first establish superiority of the experimental drug over placebo, in practicality often this key step is missed. Recently, Hida and Tango [13] considered an approach where the NI margin, δ , is defined as a pre-specified fixed difference between the treatment effects in historical placebo-controlled trial, and then tested simultaneously for both NI as well as AS hypotheses. The NI is achieved, only when one rejects both the null hypotheses jointly. They also extended their approach for binary endpoint recently [14]. A caveat of their approach is the usage of same δ for testing both the AS and NI hypothesis, which is too liberal as per FDA guideline [4]. Kwong et al. [1] proposed a modified version which takes into account the above consideration and uses different margins for NI and AS; we shall henceforth term it as MHT (Modified Hida and Tango) procedure. Ghosh et al. [15] recently proposed a Bayesian solution of the problem albeit for continuous outcome. Notably availability and usage of historical information is an integral part of the NI trial. Bayesian paradigm provides a natural path in which past trial information can be combined with the current trial to glean most possible information and thus possibly reducing sample size and other cost burden. Usage of Bayesian approaches in clinical trials and more specifically in the NI trials has a long history and can be found in the references, for e.g., [16; 17; 18; 19; 20; 21], among others. In this paper, we propose a Bayesian version of MHT procedure for binary endpoint which uses historical placebo-controlled data (or summary results) to perform Bayesian hypothesis testing in the current trial. Our approach uses direct application of Bayesian paradigm via sequential learning where prior information is combined with current trial data to improve statistical inference. Gamalo et al. [22] proposed similar approach from Bayesian point of view; however, for two-arm trials involving experimental and reference arm. Our proposed approach extends both Ghosh et al. [15] and Gamalao et al. [22], for a three-arm trial with binary end-points from Bayesian point of view. We develop methods not only for conjugate prior scenarios but also for priors having the support on the whole real line, thus allowing greater flexibility. We explore various approximations based approaches, as well as exact fully-Bayesian approaches and compare them via simulations with frequentist results. The simultaneous testing of AS and NI hypotheses is performed on simulated as well as on one published data example from mental health trial.

The rest of the article is organized as follows. In the next two sections (Section 2 and 3) we introduce a novel Bayesian methodology to design, and to perform the analysis of a three-arm NI trial for binary outcome. We consider, first, conjugate priors and then various approximations, for ease of computation. Next we consider a transformation based approach

which allows prior specification on the whole real line. We provide a detailed study of this approach under fully Bayesian setup. Section 5 presents the algorithm for simulation studies. We consider a published clinical trial dataset and analyze it by applying our proposed methodology in Section 6. We conclude the article with some discussion in Section 7.

2 Frequentist Approach for Non-Inferiority Hypothesis Testing

For a three-arm trial, from the hypothesis testing point of view and the corresponding decision rule two methods are popular; namely, 1) Fractional margin, and 2) Fixed Margin. To describe them we borrow notations from Kieser and Friede [10] and Hida and Tango [14]. We assume that primary endpoints under Experimental treatment (E), Reference treatment (R) and Placebo (P) arms, namely X_E , X_R , and X_P are independent binomially distributed as $\text{Bin}(n_E, \pi_E)$, $\text{Bin}(n_R, \pi_R)$, and $\text{Bin}(n_P, \pi_P)$ respectively. Here X_i , $i = E, R, P$ denotes the number of response in each arm and n_E , n_R , and n_P are the number of subjects assigned to each of the groups and they are not necessarily equal; π_E , π_R , and π_P represent the response probabilities ($0 \leq \pi_i \leq 1$, for $i = E, R, P$). Without loss of generality we assume that higher response probabilities indicate greater treatment benefits. Fractional margin also known as effect retention approach is popularized by Pigeot et al. [11] and Kieser and Friede [10], for continuous and binary outcomes, respectively. In this approach NI margin, Δ , is constructed as a function of treatment difference between reference and placebo arms in the current three-arm trial. One difficulty in choosing the NI margin ($\Delta > 0$) in this fashion is that the NI margin is unknown before the conduct of the current trial, where Δ is often defined as a prespecified difference between treatment and placebo in many situations. It is not clear however, how the historical information is used in the construction of the NI margin in this approach directly. In the method proposed in MHT (see Section 1), one tries to establish jointly, (i) the superiority of the reference and experimental treatments over placebo, i.e., test the null hypothesis, $K_0: \pi_R \leq \pi_P + \Delta$ (AS hypothesis) and (ii) the non-inferiority of the experimental treatment to the reference treatment i.e., test the null hypothesis $H_0: \pi_E \leq \pi_R - \Delta$ (NI hypothesis). When satisfied, these warrant two objectives, a) AS: the superiority of the R to P by more than Δ and b) NI: the non-inferiority of E with a NI margin of Δ . However, as Kwong et al. [1] reported same Δ should not be used for testing both the NI and AS, as it could lead to a too liberal test as per FDA recommendation [4]. Using the approach described in Ghosh et al. [15], one can formally write the NI testing with AS as:

$$\begin{aligned} \text{NI: } H_0: \pi_E \leq \pi_R - r\Delta \quad \text{vs.} \quad H_1: \pi_E > \pi_R - r\Delta, \quad (2.1) \\ \text{AS: } K_0: \pi_R \leq \pi_P + \Delta \quad \text{vs.} \quad K_1: \pi_R > \pi_P + \Delta. \end{aligned}$$

We want to reject both the null hypotheses H_0 and K_0 simultaneously. We term (2.1) as MHT or Modified Hida-Tango procedure. The above two alternative hypotheses H_1 and K_1 can be combined to obtain the inequality:

$$\pi_P + \Delta < \pi_R < \pi_E + r\Delta, \text{ where } 0 < r \leq 1 \text{ and } 0 \leq \pi_E, \pi_R, \pi_P \leq 1. \quad (2.2)$$

A more detailed discussion of the range of r is given in Section 3.1. A key difference of MHT approach with popular fraction margin approach is that the latter one requires simple rejection of AS hypothesis, while the former one calls for much stricter or substantial superiority of reference over placebo as specified by the AS margin. Notably in MHT approach is pre-specified and can be determined on the basis of the information from previous/historical trial(s). However, in their papers, Hida and Tango [13] and Kwong et al. [1] did not employ historical trials to construct δ , but they rather used a convenient and clinically relevant value of δ . They did mention that choice of δ should be smaller than $\frac{\delta_{past}}{2} (\delta_{past} = \pi_R^{past} - \pi_P^{past})$ as otherwise the margin will be too large for the experimental treatment to have any clinical significance, though it may still satisfy the inequality (2.2). Next, we propose a fixed margin Bayesian approach which takes into account the historical placebo-controlled trial.

3 Bayesian Approach for Non-Inferiority Testing with Binary Outcome

In this section, we discuss two Bayesian methods for non-inferiority trials for proportions. We consider construction of NI margin through the lower bound of the credible interval of the treatment effect of the active control in historical trials. In the context of two-arm NI trial that precludes placebo arm, a somewhat similar Bayesian approach has been proposed in Gamalo et al. [22]. Though both approaches are Bayesian in spirit, however quite different in terms of trial designs (two-arm vs. three-arm) and the hypotheses being tested (only NI testing vs. joint testing of NI and AS). In this article our first approach is a fully Bayesian method, however, based on the conjugate priors. We also consider an approximation based approach using normal distribution to the fully Bayesian method. The second approach is based on a transformation of the proportion parameter, with support extended on the entire real line and then putting a prior on the transformed parameter space. This “de-constraint” of the proportion parameter has several computational advantage as well as flexibility in prior specification. We also consider several approximations and a fully Bayesian version of this approach.

3.1 Bayesian Approach with Conjugate Prior

First, consider a historical placebo controlled trial for the reference treatment. There could be more than one such trial available, however we assume that those trials are combined to form a meta-analytic super-prior. There exist several available methods [21; 23] for this and we will not go into their further details as that is not the focus of this paper. Let

$X_l \sim \text{Bin}(n_{lH}, \pi_{lH})$ (for $l = R, P$) denote the random variables corresponding to historical two-arm trial and let us assume independent conjugate priors on $\pi_l \sim \text{Beta}(\alpha_{lH}, \beta_{lH})$ (for $l = R, P$).

The simple calculation shows that the posterior distributions are independent and given by,

$$\pi_l | X_{lH} \sim \text{beta}(X_{lH} + \alpha_{lH}, n_{lH} - X_{lH} + \beta_{lH}), \text{ for } l = R, P \quad (3.1)$$

respectively. The joint posterior distribution of (π_R, π_P) is the product of their marginal posterior distributions. To construct NI margin based on historical trial we need the posterior distribution of $(\pi_R - \pi_P)$ which is not available in closed form. We would like to note that Kawasaki and Miyaoka [24] and Pham-Gia and Turkkan [25] provided numerical methods to construct an HPD credible interval for the difference of binomial proportions, albeit closed form solution is impossible to obtain. However, as shown in Gamalo et al. [22], one can draw N samples from the marginal distributions of π_{RH} and π_{PH} as π_{RH}^1 ,

$\pi_{RH}^2, \dots, \pi_{RH}^N$ and $\pi_{PH}^1, \pi_{PH}^2, \dots, \pi_{PH}^N$ and then take pairwise difference to obtain sample from the distribution of the differences of $\pi_{RH} - \pi_{PH}$, which can then be sorted in increasing

magnitude as $(\pi_{RH} - \pi_{PH})^{(1)}, (\pi_{RH} - \pi_{PH})^{(2)}, \dots, (\pi_{RH} - \pi_{PH})^{(N)}$ to calculate the lower $\alpha/2$ -th

percentile point as the B . This corresponds to lower limit of the $100(1 - \alpha)\%$ credible interval for the difference $(\pi_R - \pi_P)$, which should satisfy the condition that

$P(\pi_R - \pi_P \geq \Delta^B | X_{RH}, X_{PH}) \geq 1 - \alpha/2$. The Bayesian NI margin, δ_{NI}^B , is defined as

$\delta_{NI}^B = (1 - \lambda)\Delta^B$, where λ is a fraction denoting the preservation level or desired proportion of the control effect to be retained which needs to be guided by clinical judgment. The

frequentist NI margin, δ_{NI}^F , can be constructed based on the likelihood of (X_{RH}, X_{PH}) which

is given by, $\delta_{NI}^F = (1 - \lambda)\Delta^F$, where $\Delta^F = (1 - \lambda) \left(\frac{X_{RH}}{n_{RH}} - \frac{X_{PH}}{n_{PH}} - z_{1 - \alpha/2} \sqrt{\frac{\sigma_{RH}^2}{n_{RH}^2} + \frac{\sigma_{PH}^2}{n_{PH}^2}} \right)$,

$\sigma_{lH}^2 = n_{lH} \pi_{lH} (1 - \pi_{lH})$ for $l = R, P$. However, it is more challenging to define AS margin in

the MHT context. Following the inequality (2.2), the AS margin can be defined as

$\delta_{AS}^B = \delta_{NI}^B / r = \frac{(1 - \lambda)\Delta^B}{r}$. For all practical purpose $r \in [1 - \lambda, 1]$ as when $r < (1 - \lambda)$, the AS margin asks for even stricter rejection criteria than the historical placebo controlled trial, and for $r = 1$, MHT reduces to original Hida-Tango procedure [13].

Now we consider the current three-arm trial with E, R and P arms. Since the arm E is a new intervention we consider $Beta(\alpha_E, \beta_E)$ as a prior for π_E . However, for R and P arms, since constancy is assumed for active control and placebo, we would like to use the posterior distributions of π_R and π_P in (3.1) obtained from the historical trial as their informative priors in the current trial. This informative prior setup justifies the meaningful interpretation of the NI trial. The posterior distributions for π_E , π_R and π_P are,

$$\begin{aligned} \pi_E | X_E &\sim \text{beta}(X_E + \alpha_E, n_E - X_E + \beta_E), \\ \pi_l | X_l, X_{lH} &\sim \text{beta}(X_l + X_{lH} + \alpha_{lH}, n_l + n_{lH} - X_l - X_{lH} + \beta_{lH}) \text{ for } l \in \{R, P\}. \end{aligned} \quad (3.2)$$

The posterior distribution of (π_E, π_R, π_P) is the product of their marginal posterior distributions. However, to test NI and AS simultaneously, we need the joint posterior distribution of $(\pi_E - \pi_R, \pi_R - \pi_P)$, which being the difference of beta distributions is not available in a closed form. However as indicated earlier one may obtain samples from each marginal posterior easily and, then assuming that they are independent and construct samples by taking the pair-wise differences. The test will depend upon the choice of NI and AS margins. For a specified NI margin (δ_{NI}^B) and preservation level (λ) the decision criterion is as follows. The experimental treatment is non-inferior to the active control with assay sensitivity at δ_{NI}^B/r if,

$$\begin{aligned} p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) &= P(\pi_E - \pi_R \geq -\delta_{NI}^B \cap \pi_R - \pi_P \geq \delta_{NI}^B/r | Data) \\ &= \frac{1}{M} \sum_{m=1}^M [I(\pi_E^m - \pi_R^m \geq -\delta_{NI}^B) \cap I(\pi_R^m - \pi_P^m \geq \delta_{NI}^B/r | Data)] \geq p^* \end{aligned} \quad (3.3)$$

Here, the second equation in (3.3) is based on M posterior samples, r is a fraction indicating the proportion of the NI margin to be used in the assay sensitivity, and p^* is a pre-specified value such as $p^* = 0.95, 0.975$ or any other clinically reasonable value chosen from the level (α) condition which we describe in further detail in Section 5.2. The equation (3.3) is our proposed Bayesian decision rule for jointly establishing the NI with AS.

3.2 Normal Approximation Based Approach

In this section we provide a normal approximation based solution which can avoid all the computation of previous Bayesian approach. The approximation is based on the normal approximation of beta distribution. For the historical trial, the posterior

$$\pi_l | X_{lH} \sim \text{beta}(X_{lH} + \alpha_{lH}, n_{lH} - X_{lH} + \beta_{lH}), \text{ for } l = R, P \quad (3.4)$$

is approximated by,

$$\pi_l | X_{lH} \sim \text{AN}(\mu_{lH}, \hat{\sigma}_{lH}^2), \text{ for } l = R, P, \quad (3.5)$$

where $\mu_{l^H} = \frac{X_{l^H} + \alpha_{l^H}}{n_{l^H} + \alpha_{l^H} + \beta_{l^H}}$ and $\tilde{\sigma}_{l^H}^2 = \frac{(X_{l^H} + \alpha_{l^H})(n_{l^H} - X_{l^H} + \beta_{l^H})}{(n_{l^H} + \alpha_{l^H} + \beta_{l^H})^2 (n_{l^H} + \alpha_{l^H} + \beta_{l^H} + 1)}$.

From the above, the posterior distribution of $(\pi_R - \pi_P)$ also can be shown as asymptotically normal (AN),

$$\pi_R - \pi_P | X_{R^H}, X_{P^H} \sim AN(\mu_{R^H} - \mu_{P^H}, \tilde{\sigma}_{R^H}^2 + \tilde{\sigma}_{P^H}^2). \quad (3.6)$$

A threshold value or Δ^B is obtained by solving,

$$P(\pi_{R^H} - \pi_{P^H} \geq \Delta^B | X_{R^H}, X_{P^H}) = P\left(Z \geq \frac{\Delta^B - (\mu_{R^H} - \mu_{P^H})}{\sqrt{\tilde{\sigma}_{R^H}^2 + \tilde{\sigma}_{P^H}^2}} | X_{R^H}, X_{P^H}\right) = 1 - \frac{\alpha}{2}. \quad (3.7)$$

This essentially corresponds to the lower limit of the $100(1 - \alpha)\%$ credible interval of the difference of the means $(\pi_R - \pi_P)$ which is the true effect-size. However, since it is based on the normal distribution, no computation is necessary unlike the exact Bayesian approach.

For the current trial the posterior distributions given in (3.2) are still beta distributions, which are also asymptotically normal,

$$\pi_E | X_E \sim AN(\mu_E, \tilde{\sigma}_E^2); \mu_E = \frac{X_E + \alpha_E}{n_E + \alpha_E + \beta_E}, \tilde{\sigma}_E^2 = \frac{(X_E + \alpha_E)(n_E - X_E + \beta_E)}{(n_E + \alpha_E + \beta_E)^2 (n_E + \alpha_E + \beta_E + 1)} \quad (3.8)$$

$$\pi_l | X_l, X_l^H \sim AN(\mu_l, \sigma_l^2), \text{ for } l = R, P,$$

$$\mu_l = \frac{X_l^* + \alpha_{l^H}}{n_l^* + \alpha_{l^H} + \beta_{l^H}}, \tilde{\sigma}_l^2 = \frac{(X_l^* + \alpha_{l^H})(n_l^* - X_l^* + \beta_{l^H})}{(n_l^* + \alpha_{l^H} + \beta_{l^H})^2 (n_l^* + \alpha_{l^H} + \beta_{l^H} + 1)}, \text{ where}$$

$$X_l^* = X_l + X_{l^H}, n_l^* = n_l + n_{l^H}$$

Now the posterior distributions of $(\pi_E - \pi_R)$ (for NI) and $(\pi_R - \pi_P)$ (for AS) follow bivariate normal distribution given by $AN(Y, \Sigma)$, where $Y = (\mu_E - \mu_R, \mu_R - \mu_P)$ and variance-covariance matrix is given by,

$$\Sigma = \begin{bmatrix} \tilde{\sigma}_E^2 + \tilde{\sigma}_R^2 & -\tilde{\sigma}_R^2 \\ -\tilde{\sigma}_R^2 & \tilde{\sigma}_R^2 + \tilde{\sigma}_P^2 \end{bmatrix}. \quad (3.9)$$

For a specified NI margin (δ_{NI}^B) and preservation level (λ) the decision criterion is as follows. The experimental treatment is non-inferior to the active control with assay sensitivity at δ_{NI}^B/r if,

$$\begin{aligned} p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) &= P(\pi_E - \pi_R \geq -\delta_{NI}^B \cap \pi_R - \pi_P \geq \delta_{NI}^B/r | X_E, X_R, X_P, X_{R^H}, X_{P^H}) \\ &= P\left[Z_1 \geq \frac{-\delta_{NI}^B - (\mu_E - \mu_R)}{\sqrt{\sigma_E^2 + \sigma_R^2}} \cap Z_2 \geq \frac{\delta_{NI}^B/r - (\mu_R - \mu_P)}{\sqrt{\sigma_R^2 + \sigma_P^2}}\right] \\ &\geq p^*. \end{aligned}$$

(3.10)

r is a fraction indicating the proportion of the NI margin to be used in the assay sensitivity, where Z_1 and Z_2 stand for standard normal random variables.

4 Bayesian Approach via De-Constraining Transformation

In previous section we entertained the “Beta-Binomial” conjugacy to setup our Bayesian approach. However, conjugate models are more for convenience and often lack flexibility. Also, the constrained parameter space for the response probability makes the sampling from posterior to be non-trivial. Hence, in this section, we put forward a “de-constrained” transformation approach based on *logit* transformation. The basic setup of hypothesis testing for NI and AS (see equation (2.1)) remains unchanged. We first consider, the historical trial with R and P arms, with $X_l \sim \text{Bin}(n_l, \pi_l)$ (for $l = R^H, P^H$), and let

$$\theta_l = \text{logit}(\pi_l) \Rightarrow \pi_l = \frac{e^{\theta_l}}{1 + e^{\theta_l}}, \quad -\infty < \theta_l < \infty.$$

We put a normal prior on θ_l in the hierarchical Bayesian setup as follows,

$$\begin{aligned} \theta_l | \mu_l, \sigma_l^2 &\sim N(\mu_l, \sigma_l^2) \\ \mu_l | \sigma_l^2 &\sim N(\mu_{0l}, \sigma_l^2 / \kappa_{0l}) \\ \sigma_l^2 &\sim \text{Inv} - \chi^2(\nu_{0l}, \sigma_{0l}^2) \end{aligned} \quad (4.1)$$

Where four fixed hyperparameters can be identified as location (μ_{0l}) and scale parameters ($\sigma_{0l}^2 / \kappa_{0l}$) of μ_l and degrees of freedom (ν_{0l}) and scale parameters (σ_{0l}^2) of σ_l^2 . As discussed in Ghosh et al. [19] and Gelman et al. [26], the appearance of σ_l^2 in the conditional distribution

of $\mu_l|\sigma_l^2$ act as a calibrator of the prior information about μ_l to the scale of measurement in the observed data, where $\kappa_{0l}(\geq 0)$ representing the number of prior observation on this scale. The prior sample size κ_{0l} can be adjusted with reference to the observed sample size n_l and acts as an intuitive measure characterizing the amount of prior information on μ_l . The hyper-parameters on the variance component (σ_l^2) act as a nuisance parameter and a non-informative prior can be obtained by lettering $\nu_{0l} \rightarrow 0$, resulting in Jeffreys' prior for σ_l^2 . At this point we would like to point out that Agresti and Min [27] proposed a similar prior setup for the analysis of 2×2 table, though it does not delve into fully hierarchical setup. As indicated in Agresti and Min [27], this class of prior is flexible enough to produce unimodal, bimodal, $beta(0.5, 0.5)$ as well as U-shaped distribution depending upon the specific parameter choice of σ_l^2 . It is also easy to show that the joint posterior of

$$(\mu_l, \sigma_l^2 | \theta_l) \sim N - Inv - \chi^2(\mu_{1l}, \sigma_{1l}^2 / \kappa_{1l}, \nu_{1l}, \sigma_{1l}^2), \text{ where } \mu_{1l} = \frac{\kappa_{0l}}{\kappa_{0l} + 1} \mu_{0l} + \frac{1}{\kappa_{0l} + 1} \theta_l,$$

$$\mu_{1l} = \mu_{0l} + 1, \nu_{1l} = \nu_{0l} + 1 \text{ and } \nu_{1l} \sigma_{1l}^2 = \nu_{0l} \sigma_{0l}^2 + \frac{\kappa_{0l}}{\kappa_{0l} + 1} (\theta_l - \mu_{0l})^2.$$

To construct the NI margin from historical trial data, we need to obtain sample from (θ_R, θ_P) which can be then transformed to obtain samples from (π_R, π_P) . Once such samples are obtained, one can take the pairwise difference as describe in Sub-section 3.1 to form the NI margin. For each $l \in \{R^H, P^H\}$ the sampling scheme implemented is a variation of the grouped Gibbs sampler [28; 29] which requires drawing:

1. $(\mu_l, \sigma_l^2 | \theta_l) \sim N - Inv - \chi^2(\mu_{1l}, \sigma_{1l}^2 / \kappa_{1l}, \nu_{1l}, \sigma_{1l}^2)$
2. $(\theta_l | X_l, \mu_l, \sigma_l^2) \propto \text{Binomial-Likelihood}(\theta_l | X_l) \times N(\theta_l | \mu_l, \sigma_l^2) = \frac{e^{X_l \theta_l}}{(1 + e^{\theta_l})^{n_l}} \times N(\theta_l | \mu_l, \sigma_l^2)$

The above implementation is computationally feasible with acceptable convergence rate. For θ_l we adopt Acceptance-Rejection algorithm as it is easy to show

$$M_{\max} = \sup_{\theta_l} \frac{\text{Posterior}(\theta_l | X_l, \mu_l, \sigma_l^2)}{N(\theta_l | \mu_l, \sigma_l^2)} = \log\left(\frac{X_l}{n_l - X_l}\right). \text{ This greatly facilitate the choice of candidate}$$

density as normal distribution with satisfactory acceptance probability. Next, we describe an approximation based approach, followed by an exact approach, under de-constrained transformation.

4.1 Beta-Approximation via Moment Matching

One disadvantage of the previously discussed approach is the unavailability of the distribution of π_l for $l \in \{R^H, P^H\}$ in a closed form. As mentioned in Section 3.1 our main idea under Bayesian setup is to put the historical posterior as the prior for R and P arm, albeit if it is available in a tractable form. To alleviate this problem, we explored a moment matching approach via delta method. Using delta method it is easy to approximate the prior for mean and variance of π_l as,

$$\pi_l \simeq F(\mu_l) + (\theta_l - \mu_l)F'(\mu_l) \text{ where } F(x) = \frac{e^x}{1 + e^x} \Rightarrow E[\pi_l] \simeq \frac{1}{1 + e^{-\mu_l}}, V[\pi_l] \simeq \frac{\sigma_l^2 e^{-2\mu_l}}{(1 + e^{-\mu_l})^4}. \quad (4.2)$$

Next we approximate the historical prior of π_l for $l \in \{R^H, P^H\}$ via beta distribution as

$$\pi_l \sim \text{beta}(K_l V_l, K_l(1 - V_l)), \text{ for } l = R^H, P^H \text{ where } V_l = \frac{1}{1 + e^{-\mu_l}} \text{ and } \frac{V_l(1 - V_l)}{K_l + 1} \quad (4.3)$$

$$= \frac{\sigma_l^2 e^{-2\mu_l}}{(1 + e^{-\mu_l})^4}.$$

From the above equation one can solve (K_l, V_l) in terms of (μ_l, σ_l^2) . Since the prior is approximated in beta-conjugate form, one can now easily apply the method presented in Section 3.1 to carry out the Bayesian decision. However, one disadvantage of this approximation is that since we approximate the normal prior via beta, the flexibility of de-constraining the parameter is compromised. This method also shows a situation, when there is a substantial prior information available from experts about R and P arms, which is not necessarily constrained in the interval $[0,1]$.

4.2 An Exact Sampling based Approach

In previous two sub-sections we described how to use de-constrained approach to put normal prior on the transformed parameter space (θ) . We have also described a sampling based approach to construct NI margin (). However our Bayesian idea of placing posteriors of R and P arms from historical trials, as priors for the current trial is challenging, as we never evaluated the joint posterior of (π_R, π_P) . In previous Subsection we propose to solve this problem via moment based Beta-approximation. In this section we explored a sampling based exact approach, however to do so in the current trial we model the transformed parameters. So for E, R and P arms, with $X_l \sim \text{Bin}(n_l, \pi_l)$ (for $l = E, R, P$), we let again

$$\theta_l = \text{logit}(\pi_l) \Rightarrow \pi_l = \frac{e^{\theta_l}}{1 + e^{\theta_l}}, -\infty \leq \theta_l \leq \infty.$$

For R and P arms the posteriors, in the historical trials are given by,

$$(\theta_l | X_{l^H}, \mu_{l^H}, \sigma_{l^H}^2) \propto \frac{e^{X_{l^H} \theta_l}}{(1 + e^{\theta_l})^{n_{l^H}}} \times N(\theta_l | \mu_{l^H}, \sigma_{l^H}^2)$$

Since all distributions are proper the posterior propriety is satisfied, however nice closed form solution cannot be obtained. Next, we plan to use above proper posteriors as priors in the current trial. For E arm, since there is no historical information available, a vague prior is proposed which is similar to (4.1). This leads to posterior distributions in the current trial as

$$\begin{aligned} (\theta_E | X_E, \mu_E, \sigma_E^2) &\sim \frac{e^{X_E \theta_E}}{(1 + e^{\theta_E})^{n_E}} \times N(\theta_E | \mu_E, \sigma_E^2), \\ (\theta_l | X_l, X_{l^H}, \mu_{l^H}, \sigma_{l^H}^2) &\sim \frac{e^{X_l \theta_l + X_{l^H} \theta_l}}{(1 + e^{\theta_E})^{n_l + n_{l^H}}} \times N(\theta_l | \mu_{l^H}, \sigma_{l^H}^2), \text{ for } l \in \{R, P\}. \end{aligned} \quad (4.4)$$

Though not explicitly shown, the hyperpriors are all embedded in the above posteriors as specified in (4.1). Since we cannot draw samples from the posterior distribution of (π_E, π_R, π_P) directly, we first obtain samples from $(\theta_E, \theta_R, \theta_P)$, and then use inverse transformation. Using M draws we compute,

$$\begin{aligned} p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) &= P(\pi_E - \pi_R \geq -\delta_{NI}^B \cap \pi_R - \pi_P \geq \delta_{NI}^B / r | \text{Data}) \\ &= \frac{1}{M} \sum_{m=1}^M [I(\pi_E^m - \pi_R^m \geq -\delta_{NI}^B) \cap I(\pi_R^m - \pi_P^m \geq \delta_{NI}^B / r | \text{Data})] \geq p^*, \end{aligned} \quad (4.5)$$

where p^* is pre-specified, such as $p^* = 0.975, 0.95$ or any other clinically meaningful values. In the simulation study, we shall use 0.975. We refer to equation (4.5) as the proposed Bayesian decision criterion.

5 Simulation Study

In this section, we enumerate a few simulation studies to benchmark the performance of the above Bayesian procedures in comparison to the frequentist approach. Numerical results are then presented in Tables 1–6.

5.1 Simulation Steps

The following Markov chain Monte Carlo (MCMC) simulation steps are used to calculate the power and type-I error for both beta-binomial as well as de-constrained normal prior case. A Beta(1,1) prior is assumed for π_b , $I \in \{R^H, P^H\}$ and π_E under Beta-Binomial setup, while under de-constrained hierarchical setup $\sigma_I^2 \sim Inv - \chi^2(4, 1)$ and $\mu_I | \sigma_I^2 \sim N(0, \sigma_I^2)$.

1. Specify $n_{P^H}, n_{R^H}, \pi_{P^H}, \pi_{R^H}, n_P, n_R, n_E, \pi_P, \pi_R, \pi_E$, where $\pi_P < \pi_R$ in order to generate X_P^H, X_R^H, X_P, X_R and X_E .
2. Generate the data from $X_I^H \sim Bin(\pi_{I^H}, n_{I^H})$, for $I \in \{P, R\}$.
3. Generate NMCMC samples on π_b , $I \in \{P, R\}$ from the historical posterior given in (3.1).
4. Calculate the difference $(\pi_{R^H} - \pi_{P^H})$ for N pairs and sort them in increasing order to calculate the Bayesian lower $\alpha/2$ -th percentile point as B .
5. Let π_E be less than $\pi_R - \delta$ (for calculating the type-I error) and generate data $X_I \sim Bin(n_I, \pi_I)$ for $I \in \{P, R, E\}$. Note that for a specified preservation level λ , $\delta_{NI}^B = (1 - \lambda)\Delta^B$ for the Bayesian setting, and $\delta_{NI}^F = (1 - \lambda)\Delta^F$ for the frequentist setting.
6. From the current trial generate MMCMC samples using the conditional posteriors specified in (3.2).
7. Compute the samples on $(\pi_E - \pi_R)$ and $(\pi_R - \pi_P)$ for M pairs in Step 6.
8. Calculate the posterior probability

$$p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) = \frac{1}{M} \sum_{m=1}^M \left[I(\pi_E^m - \pi_R^m \geq -\delta_{NI}^B) \cap I(\pi_R^m - \pi_P^m \geq \delta_{NI}^B / r | X_P, X_E, X_R, X_P^H, X_R^H) \right]$$
9. Bayesian criterion: If $p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) \geq p^*$, increase the COUNTS by 1; otherwise 0.
10. Go back to Step 2 and repeat the simulation n^* times. Calculate the power by using the COUNTS divided by n^* , where $n^* (10,000)$ is a large number chosen a priori.

Note that for the frequentist procedure, Step 9 needs to be replaced by the frequentist decision criteria such as the one based on comparison of p-values with α . For normal approximation as well as for de-constrain transformation the priors as well as the posteriors in Steps 4–7 need appropriate replacement in the above algorithm. Also note that since both NI and AS hypotheses are tested jointly additional care should be taken. For example Kwong et al.[1] proposed controlling the family-wise error rate (FWER), however, as indicated in Ghosh et al. [15] under Intersection-Union (IU) setup such adjustments are not necessary for controlling of type-I error. As shown in Berger [30], under IU testing setup,

each individual sub-hypothesis (e.g., NI and AS in the current context) can be tested at α level, which is also regulatory position regardless of the number of co-primary hypotheses. The resulting All-or-None test is referred as MIN test (Laska and Meisner [31]) which requires no level adjustments, however it tends to be very conservative. As a result some adjustments to the level is recommended by many researchers, for example, Relaxed IU test by Deng et al. [32] and Average Testing Error by Chuang et al. [33]. One of the major points of these approaches is that when test-statistics of the co-primary hypotheses are moderately correlated (with correlation < 0.2 , see equation (3.9) for present case) and sample size for each arm is large enough (> 75), the gain of adjusting/relaxing type-I error to the power is minimal. Since NI trials typically enroll bigger sample size compared to superiority trials, we did not proceed further in modifying the definition of type-I error. Nevertheless, some recent work on using ATE and its ramification on the trial design is worthwhile to note [34] and we think more work in this direction is necessary from frequentist perspective.

5.2 Simulation Results

First, we determine the number of MCMC samples required to have stable results for conducting simulation studies for the empirical type-I error and power of the proposed Bayesian methods. A trace plot of posterior estimates for each parameter suggests $N = M = 50,000$ MCMC samplers, with 500 burn-ins, are more than sufficient to produce stable estimates of the parameters. Throughout the simulation study, for historical trial, we assume Beta(1,1) as a prior for both R^H and P^H arms and consider the following specification of the parameters: $\pi_{P^H} = 0.2$, $\pi_{R^H} = 0.75$. For our entire the simulation study we set,

$\pi_E = \pi_R - \delta_{NI}^B + \xi_1$ and $\pi_P = \pi_R - \delta_{NI}^B/r - \xi_2$, where ξ_1 and ξ_2 are non-negative constants. It is possible to consider different values of ξ_1 and ξ_2 with the restriction that $0 \leq \pi_P \leq \pi_E \leq 1$ and we choose $r = 1 - \lambda$. This specific choice of r is the most conservative AS margin obtained from the historical two-arm trial as discussed in Ghosh et al. [15]. If it clinically makes sense, it is possible to choose other values of $r \in [1 - \lambda, 1]$ which will make the AS margins less wider resulting in less conservative decision procedures. The non-inferiority of a new treatment is established by rejecting both H_0 and K_0 simultaneously. Therefore, the hypothesis tests in (2.1) is the Intersection-Union test. Hence, the the type-I errors for Bayesian methods are evaluated as $\Pr(\text{Decision Region for non-inferiority Goal} | H_0 \cup K_0)$. Another important criterion is the choice of p^* . Following the frequentist setup, we choose $p^* = 0.975$; however as reported in Gamalo et al. [22] this specific choice gives the type-I error that is too restrictive. One way to alleviate this problem is performing Bayesian calibration. The idea is to select a threshold for significance that achieves desired frequentist's property (e.g. type-I error $= \alpha$) in the Bayesian testing (Guo and Heitjan (2010), Weiss (1997)). We performed Bayesian calibration for different choices of λ which is presented in Tables 1,3, and 5 for different simulation scenarios. For the calibrated result for a specific λ , we resort to that specific p^* which gives sufficient control over the type-I error, without being too conservative under each setup. Note that to compare our results with the frequentist approach, we also reported results for fixed $p^* = 0.975$. To compute the type-I, we set $\xi = 0$, while for the power analysis we entertained a wide range of values. Our simulations results are reported for both sample and large sample sizes. For the small sample

case, we consider $n_{P^H} = n_{R^H} = 30$ for historical trial and $n_P = n_R = n_E = 30$ for the current trial. For the large sample case, we set $n_{P^H} = n_{R^H} = 300$ and for the current trial we consider two different sample sizes ($n_P = n_R = n_E = 100$ and $n_P = n_R = n_E = 300$).

Under both sample and large sample cases, we consider three different scenarios with respect to the constancy assumption of the reference treatment in comparison to the historical and current trial. Scenario 1: When the effect of historical reference is same as the effect in current trial; i.e., $\pi_{R^H} = \pi_R$ and the constancy assumption is satisfied. Scenario 2:

When the effect of historical reference is slightly superior compared to the effect of current trial; i.e., $\pi_{R^H} > \pi_R$. Scenario 3: When the effect of historical reference is slightly inferior

compared to the effect of current trial; i.e., $\pi_{R^H} < \pi_R$. We consider $\pi_{R^H} = 0.75$ and $\pi_{P^H} = 0.2$

for all scenarios. All the numerical results reported in this paper are obtained based $n^* = 10,000$ simulations. Table 1 shows the type-I error which is estimated as the number of rejections out of $n^* = 10,000$. For fixed p^* the un-calibrated Bayesian methods (e.g., UCBM, UCPA and UCBD) are too conservative. Calibrated p^* is obtained so that type-I error is less than < 0.025 for all λ without being too conservative by relaxing the p^* to a lower value. To compute the power we varied ξ from 0.05 to 0.2. Under both sample and large sample cases, the power for calibrated Bayesian methods are more. In fact the method based on De-Constrained prior has the highest power followed by calibrated Bayesian method based on Beta prior as seen in Tables 2, 4 and 6. However when they are not calibrated all three Bayesian methods are more powerful than the frequentist method but the power is quite low for small ξ . Among three simulation scenarios power is lowest when $\pi_{R^H} > \pi_R$ and highest when $\pi_{R^H} < \pi_R$. A similar observation is also made in by Gamalo et al.

[22], albeit for two-arm trial. The normal approximation to posterior distribution works somewhat close to the frequentist's method and the approximation is pretty close to the exact Bayesian method for high values of ξ .

6 Application

We applied our developed Bayesian and frequentist's methods on a published dataset. The dataset is coming from a depression trial described in Higuchi et al. [35]. This specific study was also considered in the paper by Hida and Tango [13; 14] and Ghosh et al. [15]. In the first two papers a frequentist's version of the problem is proposed first for continuous outcome, and then for a modified binary outcome which is directly related to the current paper. The primary objective of the study is to compare the efficacy and safety of duloxetine (E) with those of paroxetine (R) and placebo (P). This is a double-blinded, randomized, active-controlled, parallel-group study of a 6-week treatment with duloxetine ($n_E = 147$) or paroxetine ($n_R = 148$) or placebo ($n_P = 145$). The primary endpoint is the change in HAMD-17 total score at the end of 6-week, which is continuous type. For their frequentist exposition Hida and Tango [14] only entertained two binary outcomes namely: Remission and Response. Remission is defined as maintaining HAMD-17 score to be ≤ 7 at 6 weeks, which is a secondary end-point of the study. Response is defined as the reduction of more

than 50% total HAMD-17 score at 6 weeks. The basic data is presented in their article in terms of Response and Remission and also repeated in Table 7. For our illustration we did two separate analysis with these two outcomes to compare various Bayesian and frequentist methods. Non-informative priors are used in all setup to begin with, and to compute NI and AS margins. We used calibrated value of $p^* = 0.768$ under the constancy assumption $(\pi_{RH} = \pi_R)$ for Bayesian methods, and used $p^* = 0.975$ for the frequentist method.

We report our result in Table 8. We also report the average posterior probability for simultaneous rejection of H_0 and K_0 for different Bayesian margins, which can be compared with p^* to deduce the Bayesian decision. Since non-informative priors are used in the historical trial, both the Bayesian (especially under conjugate Beta-Binomial) and frequentist margins are very close. Table 8 used calibration which also results in not too conservative type-I error (< 0.025). Under no calibration, the type-I error < 0.001 tends to be extremely conservative as showed in the simulation section. We would like to point out that under non-calibration, i.e., if $p^* = 0.975$ is used Bayesian and frequentist methods yield identical decision rule. Also note that, since we have used $r = 1 - \lambda$ throughout for the AS margin, it is not affected by changing the λ values. Hida and Tango [14] reported lack of assay sensitivity resulting rejection of NI which is also evident from “Frequentist Decision” in our result. However using Bayesian rule, with De-Constrained prior NI with AS is established. This is also evident from improved posterior probability estimate. The difference in decision, in testing AS hypothesis, between the Bayesian and frequentist approaches becomes clear when we compare the 95% confidence interval of $(\pi_R - \pi_P)$ reported as [0.0021, 0.2045] with a pre-specified $\delta = 0.1$ (see Hida and Tango [14]), while the corresponding Bayesian credible interval for the same reported as [0.070, 0.203] in Table 9, is considerably narrower. The result for outcome “Response” is reported in Tables 10 and 11. Hida and Tango [14] reported NI but lack of AS. Under Bayesian setup the result is reversed for all $\lambda < 0.7$. This difference in decision is clear when we see Bayesian credible interval in Table 11.

7 Discussion

Non-inferiority trials are unique in their setup and availability of relevant historical information which may greatly influence the ultimate success of the trial. Bayesian paradigm provides an unique approach allowing combination of prior and current trial data in a single framework. The challenges of Bayesian approach are its subjectivity in prior specification as well as resulting computational complication when posterior is not available in closed form. In this paper, we have developed Bayesian approach for testing non-inferiority under various choice of prior distribution. We explored standard conjugate prior and also specified more flexible choices of priors under transformation of model parameter. We explored various approximation which can avoid lengthy computation albeit at the cost of slight loss of precision. We observed that with the Bayesian calibration one can control type-I error, with significant gain in the power. Albeit we should also mention that this power gain is only significant for moderate to large value of ξ which essentially controls the “effect size”.

We also applied both the Bayesian and frequentist methods on a clinical trial data. The results suggest that the Bayesian methods perform favorably in all situations, and that these methods do not depend on any asymptotic approximation as frequentist method. Notably, with binary end-point risk difference is not the only functional of interest. One may frame NI and AS hypothesis in terms of log odds and/or risk ratio. While some frequentist work exists for two-arm NI trial almost no such method exists for those endpoints for a three-arm trial. We are currently developing both frequentist and Bayesian methods for these types of functional. Another interesting extension could be semi or non-parametric extension of our approach. Limited work has been done for a two-arm trial, but much need to be explored for a three-arm trial, which is also one of our future research interest.

Acknowledgement

The research of last/corresponding author is partly supported by PCORI contract number ME-1409-21410 and NIH grant number P30-ES020957.

References

- [1]. Kwong KS, Cheung SH, Hayter AJ, and Wen MJ. Extension of three-arm non-inferiority studies to trials with multiple new treatments. *Statistics in Medicine*, 31(24):2833–2843, 2012. [PubMed: 22744965]
- [2]. D’Agostino RB, Massaro JM, and Sullivan LM. Noninferiority trials: Design concepts and issues—the encounters of academic consultants in statistics. *Statistics in Medicine*, 22(2):169–186, 2003. [PubMed: 12520555]
- [3]. FDA. Non-Inferiority Clinical Trials to Establish Effectiveness: Guidance for Industry, 2016.
- [4]. FDA. Food and Drug Administration Guidance for Industry Noninferiority Clinical Trials, 2010.
- [5]. ICHE9. ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use., 2009.
- [6]. ICHE10. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use., 2009.
- [7]. EMA. Guideline on the choice of the noninferiority margin (Doc. Ref. EMEA/CPMP/EWP/215), 2005.
- [8]. Hung HMJ and Wang SJ. Multiple testing of noninferiority hypotheses in active controlled trials. *Journal of Biopharmaceutical Statistics*, 14(2):327–335, 2004. [PubMed: 15206530]
- [9]. Schumi J and Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*, 12(2): 106–118, 2011. [PubMed: 21539749]
- [10]. Kieser M and Friede T. Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in medicine*, 26(2):253–273, 2007. [PubMed: 16685703]
- [11]. Pigeot I, Schfer J, Rhmel J, and Hauschke D. Assessing noninferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22(6):883–899, 2003. [PubMed: 12627407]
- [12]. Koch A and Rohmel J. Hypothesis testing in the gold standard design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of biopharmaceutical statistics*, 14(2):315–325, 2004. [PubMed: 15206529]
- [13]. Hida E and Tango T. On the three-arm noninferiority trial including a placebo with a prespecified margin. *Statistics in Medicine*, 30(3):224–231, 2011. [PubMed: 21213340]
- [14]. Hida E and Tango T. Three-arm noninferiority trials with a prespecified margin for inference of the difference in the proportions of binary endpoints. *Journal of biopharmaceutical statistics*, 23(4):774–789, 2013. [PubMed: 23799810]

- [15]. Ghosh S, Ghosh S, and Tiwari RC. Bayesian approach for assessing non-inferiority in a three-arm trial with pre-specified margin. *Statistics in medicine*, 2015.
- [16]. Simon R Bayesian design and analysis of active control clinical trials. *Biometrics*, 55(2):484–487, 1999. [PubMed: 11318204]
- [17]. Spiegelhalter DJ, Abrams KR, and Myles JP. Bayesian approaches to clinical trials and healthcare evaluation. New York: Wiley, 2004.
- [18]. Liu J, Hsueh H, and Hsiao C. A bayesian non-inferiority approach to evaluation of bridging studies. *Journal of Biopharmaceutical Statistics*, 14(2):291–300, 2004. [PubMed: 15206527]
- [19]. Ghosh P, Nathoo F, Gnen M, and Tiwari RC. Assessing noninferiority in a three-arm trial using the bayesian approach. *Statistics in Medicine*, 30(15):1795–1808, 2011. [PubMed: 21520456]
- [20]. Gamalo MA, Wu R, and Tiwari RC. Bayesian approach to non-inferiority trials for normal means. *Statistical Methods in Medical Research*, Published online 21 5 2012.
- [21]. Gamalo MA, Tiwari RC, and LaVange LM. Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. *Pharmaceutical Statistics*, 13(1):25–40, 2014. [PubMed: 23913880]
- [22]. Gamalo MA, Wu R, and Tiwari RC. Bayesian approach to noninferiority trials for proportions. *Journal of biopharmaceutical statistics*, 21(5):902–919, 2011. [PubMed: 21830922]
- [23]. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, and Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014. [PubMed: 25355546]
- [24]. Kawasaki Youhei and Miyaoka Etsuo. A posterior density for the difference between two binominal proportions and the highest posterior density credible interval. *Journal of the Japan Statistical Society*, 40(2):265–275, 2010.
- [25]. Turkkan Noyan and Pham-Gia T. Computation of the highest posterior density interval in bayesian analysis. *Journal of statistical computation and simulation*, 44(3–4):243–250, 1993.
- [26]. Gelman Andrew, Carlin John B, Stern Hal S, and Rubin Donald B. Bayesian data analysis, volume 2 Taylor & Francis, 2014.
- [27]. Agresti A and Min Y. Frequentist performance of bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics*, 61(2):515–523, 2005. [PubMed: 16011699]
- [28]. Geman Stuart and Geman Donald. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.
- [29]. Gelfand Alan E and Smith Adrian FM. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.
- [30]. Berger RL. Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, 24(4): 295–300, 1982.
- [31]. Laska EM and Meisner MJ. Testing whether an identified treatment is best. *Biometrics*, pages 1139–1151, 1989. [PubMed: 2611321]
- [32]. Deng X, Xu J, and Wang C. Improving the power for detecting overlapping genes from multiple dna microarray-derived gene lists. *Statistics in Medicine*, 9(Suppl 6):S14, 2008.
- [33]. Chuang-Stein Christy, Stryszak Paul, Dmitrienko Alex, and Offen Walter. Challenge of multiple co-primary endpoints: a new approach. *Statistics in medicine*, 26(6):1181–1192, 2007. [PubMed: 16927251]
- [34]. Kieser Meinhard and Stucke Kathrin. Assessing additional benefit in noninferiority trials. *Biometrical Journal*, 2016.
- [35]. Higuchi T, Murasaki M, and Kamijima K. Clinical evaluation of duloxetine in the treatment of major depressive disorder-placebo and paroxetine-controlled double-blinded comparative study. *Japanese Journ of Clinical Psychopharmacology*, 12:1613–1634, 2009.

Table 1:

Simulation study 1: The simulated type-I errors using $n_{p^H} = n_{R^H} = n_R = n_P = n_E = 300$,

$\pi_{p^H} = 0.2$, $\pi_{R^H} = \pi_R = 0.75$; CBM: Calibrated Bayesian Method based on Beta Prior; CPA: Calibrated

Bayesian Method based on Posterior Approximation via Normal (section 3.2); CBD: Calibrated Bayesian De-Constraining prior. CBD approach is based on section 4.2, “U” stands for uncalibrated type-I error for each of the above method at $p^* = 0.975$.

λ	p^*	CBM	CPA	CBD	p^*	UCBM	UCPA	UCBD
0.0	0.757	0.0230	0.0090	0.016	0.975	0.0008	0.0001	0.0003
0.25	0.763	0.0228	0.0156	0.018	0.975	0.0008	0.0001	0.0005
0.50	0.765	0.0212	0.0130	0.020	0.975	0.0006	0.0002	0.0003
0.75	0.768	0.0226	0.0182	0.019	0.975	0.0006	0.0003	0.0004

Table 2:

Power for Simulation study 1 ($\pi_{RH} = \pi_R$): Empirical power is computed for $p^* = 0.768$ for calibrated Bayesian methods. For Un-calibrated and Frequentist's methods $p^* = 0.975$.

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
$n_E = n_R = n_P = 30$								
0.0	0.05	0.189	0.175	0.182	0.003	0.011	0.003	$4e^{-04}$
0.0	0.10	0.494	0.395	0.450	0.033	0.024	0.030	0.011
0.0	0.15	0.742	0.651	0.729	0.146	0.144	0.147	0.074
0.0	0.20	0.903	0.854	0.892	0.405	0.420	0.409	0.302
0.25	0.05	0.177	0.152	0.171	0.002	0.009	0.003	$6e^{-04}$
0.25	0.10	0.444	0.382	0.425	0.027	0.027	0.025	0.0134
0.25	0.15	0.714	0.645	0.719	0.127	0.132	0.132	0.0882
0.25	0.20	0.874	0.856	0.892	0.387	0.409	0.400	0.320
0.50	0.05	0.171	0.149	0.160	0.003	0.008	0.002	0.0012
0.50	0.10	0.423	0.378	0.443	0.023	0.037	0.033	0.0114
0.50	0.15	0.711	0.651	0.713	0.120	0.147	0.136	0.091
0.50	0.20	0.891	0.878	0.893	0.379	0.387	0.397	0.343
0.75	0.05	0.164	0.146	0.170	0.001	0.009	0.021	0.002
0.75	0.10	0.421	0.390	0.444	0.024	0.030	0.045	0.014
0.75	0.15	0.734	0.687	0.730	0.128	0.112	0.128	0.101
0.75	0.20	0.921	0.903	0.930	0.418	0.405	0.412	0.401
$n_E = n_R = n_P = 100$								
0.0	0.05	0.443	0.418	0.450	0.026	0.021	0.022	0.001
0.0	0.10	0.818	0.790	0.823	0.265	0.258	0.277	0.067
0.0	0.15	0.976	0.969	0.980	0.728	0.731	0.730	0.463
0.0	0.20	0.999	0.998	1.000	0.952	0.948	0.965	0.842
0.25	0.05	0.404	0.375	0.420	0.027	0.026	0.020	0.002
0.25	0.10	0.787	0.772	0.800	0.224	0.227	0.230	0.072
0.25	0.15	0.967	0.955	0.970	0.642	0.656	0.685	0.429
0.25	0.20	0.997	0.993	0.999	0.937	0.936	0.934	0.842
0.50	0.05	0.399	0.370	0.399	0.021	0.025	0.023	0.002
0.50	0.10	0.786	0.769	0.793	0.234	0.231	0.211	0.071
0.50	0.15	0.969	0.952	0.966	0.677	0.670	0.660	0.451
0.50	0.20	0.997	0.997	0.999	0.940	0.944	0.936	0.871
0.75	0.05	0.380	0.375	0.395	0.018	0.025	0.019	0.004
0.75	0.10	0.803	0.785	0.807	0.235	0.256	0.234	0.081
0.75	0.15	0.977	0.968	0.981	0.738	0.745	0.738	0.510

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
0.75	0.20	0.999	0.999	0.999	0.975	0.976	0.973	0.912
$n_E = n_R = n_P = 300$								
0.0	0.05	0.675	0.662	0.667	0.108	0.091	0.105	0.031
0.0	0.10	0.986	0.966	0.987	0.796	0.790	0.801	0.661
0.0	0.15	1.000	0.999	1.000	0.991	0.994	0.999	0.984
0.0	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.05	0.634	0.625	0.620	0.085	0.081	0.074	0.033
0.25	0.10	0.977	0.961	0.978	0.761	0.748	0.732	0.637
0.25	0.15	1.000	0.995	1.000	0.989	0.991	0.992	0.974
0.25	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	0.05	0.628	0.617	0.626	0.078	0.086	0.075	0.033
0.50	0.10	0.978	0.969	0.981	0.747	0.737	0.742	0.633
0.50	0.15	1.000	1.000	1.000	0.992	0.991	0.990	0.980
0.50	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.05	0.626	0.613	0.636	0.084	0.088	0.083	0.031
0.75	0.10	0.984	0.977	0.987	0.776	0.785	0.786	0.672
0.75	0.15	1.000	1.000	1.000	0.995	0.997	0.996	0.989
0.75	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3:

Simulation study 2: The simulated type-I error is obtained using $n_{P^H} = n_{R^H} = n_R = n_P = n_E = 300$, $\pi_{P^H} = 0.2$,

$\pi_{R^H} = 0.75 > \pi_R = 0.7$.

λ	p^*	CBM	CPA	CBD	p^*	UCBM	UCPA	UCBD
0.00	0.630	0.023	0.015	0.017	0.975	0.0001	0.0003	0.0002
0.25	0.667	0.024	0.018	0.020	0.975	0.0002	0.00001	0.0001
0.50	0.658	0.021	0.019	0.021	0.975	0.0001	0.0002	0.0003
0.75	0.660	0.020	0.016	0.020	0.975	0.0001	0.0002	0.0002

Table 4:

Power for Simulation study 2 $\left(\pi_{RH} > \pi_R\right)$: Empirical power computed for $p^* = 0.667$ for calibrated Bayesian methods; For Un-calibrated and Frequentist's method $p^* = 0.975$.

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
$n_E = n_R = n_P = 30$								
0.0	0.05	0.206	0.208	0.210	0.003	0.0048	0.003	$2e^{-04}$
0.0	0.10	0.590	0.575	0.589	0.025	0.035	0.030	0.012
0.0	0.15	0.774	0.760	0.791	0.128	0.125	0.147	0.095
0.0	0.20	0.912	0.908	0.911	0.358	0.339	0.409	0.202
0.25	0.05	0.231	0.212	0.228	0.004	0.004	0.003	$6e^{-04}$
0.25	0.10	0.541	0.533	0.549	0.020	0.032	0.025	0.012
0.25	0.15	0.798	0.751	0.810	0.138	0.129	0.132	0.0982
0.25	0.20	0.912	0.916	0.931	0.412	0.397	0.400	0.210
0.50	0.05	0.210	0.205	0.201	0.004	0.003	0.002	$2e^{-04}$
0.50	0.10	0.547	0.550	0.572	0.031	0.031	0.033	0.013
0.50	0.15	0.809	0.753	0.806	0.125	0.122	0.136	0.095
0.50	0.20	0.917	0.907	0.923	0.390	0.393	0.397	0.243
0.75	0.05	0.206	0.207	0.201	0.003	0.003	0.021	$2e^{-04}$
0.75	0.10	0.553	0.541	0.567	0.048	0.032	0.045	0.014
0.75	0.15	0.777	0.764	0.797	0.128	0.125	0.128	0.111
0.75	0.20	0.927	0.920	0.930	0.408	0.390	0.412	0.209
$n_E = n_R = n_P = 100$								
0.0	0.05	0.256	0.228	0.243	0.003	0.002	0.002	0.0002
0.0	0.10	0.713	0.679	0.682	0.096	0.072	0.079	0.021
0.0	0.15	0.948	0.920	0.939	0.449	0.445	0.455	0.208
0.0	0.20	0.996	0.956	1.000	0.790	0.773	0.796	0.589
0.25	0.05	0.259	0.225	0.242	0.003	0.003	0.003	0.0004
0.25	0.10	0.649	0.605	0.643	0.075	0.073	0.064	0.025
0.25	0.15	0.921	0.915	0.927	0.381	0.389	0.389	0.207
0.25	0.20	0.990	0.950	0.987	0.732	0.719	0.732	0.585
0.50	0.05	0.248	0.227	0.242	0.003	0.003	0.004	0.0004
0.50	0.10	0.642	0.612	0.636	0.059	0.070	0.061	0.022
0.50	0.15	0.910	0.899	0.916	0.376	0.374	0.368	0.193
0.50	0.20	0.993	0.960	0.985	0.756	0.731	0.752	0.598
0.75	0.05	0.253	0.234	0.248	0.005	0.004	0.005	0.0002
0.75	0.10	0.644	0.609	0.636	0.067	0.078	0.074	0.027

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
0.75	0.15	0.925	0.905	0.926	0.387	0.414	0.392	0.219
0.75	0.20	0.997	0.958	0.999	0.791	0.795	0.795	0.636
$n_E = n_R = n_P = 300$								
0.0	0.05	0.462	0.431	0.460	0.013	0.010	0.012	0.007
0.0	0.10	0.957	0.920	0.957	0.507	0.502	0.507	0.344
0.0	0.15	0.999	0.950	1.000	0.951	0.958	0.947	0.861
0.0	0.20	1.000	1.000	1.000	0.975	1.000	0.985	0.989
0.25	0.05	0.423	0.409	0.439	0.012	0.015	0.009	0.009
0.25	0.10	0.942	0.915	0.938	0.440	0.456	0.459	0.332
0.25	0.15	1.000	0.945	0.999	0.933	0.939	0.931	0.854
0.25	0.20	1.000	0.990	1.000	0.982	1.000	0.983	0.988
0.50	0.05	0.440	0.415	0.440	0.008	0.011	0.014	0.008
0.50	0.10	0.938	0.908	0.933	0.431	0.438	0.430	0.318
0.50	0.15	1.000	0.950	0.999	0.927	0.930	0.932	0.854
0.50	0.20	1.000	1.000	1.000	0.979	0.999	0.983	0.989
0.75	0.05	0.426	0.409	0.436	0.012	0.017	0.013	0.007
0.75	0.10	0.942	0.911	0.944	0.440	0.460	0.445	0.340
0.75	0.15	1.000	0.960	0.999	0.951	0.946	0.946	0.862
0.75	0.20	1.000	1.000	1.000	0.983	1.000	0.980	0.989

Table 5:

Simulation study 3: The simulated type-I error is obtained using $n_{P^H} = n_{R^H} = n_R = n_P = n_E = 300$, $\pi_{P^H} = 0.2$,

$\pi_{R^H} = 0.75 < \pi_R = 0.8$.

λ	p^*	CBM	CPA	CBD	p^*	UCBM	UCPA	UCBD
0.00	0.866	0.0240	0.0208	0.0210	0.975	0.0002	0.0005	0.0020
0.25	0.866	0.0208	0.0238	0.0190	0.975	0.0002	0.0004	0.0016
0.50	0.863	0.0235	0.0188	0.0200	0.975	0.0004	0.0002	0.0013
0.75	0.863	0.0202	0.0166	0.0190	0.975	0.0006	0.0006	0.0010

Table 6:

Power for Simulation study 3 $\left(\pi_{RH} < \pi_R\right)$; Empirical power is computed for $p^* = 0.866$ for calibrated Bayesian methods; For Un-calibrated and Frequentist's method $p^* = 0.975$.

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
$n_E = n_R = n_P = 30$								
0.0	0.05	0.230	0.235	0.251	0.008	0.019	0.0210	0.008
0.0	0.10	0.500	0.486	0.587	0.100	0.092	0.110	0.091
0.0	0.15	0.807	0.744	0.839	0.335	0.328	0.351	0.284
0.0	0.20	0.922	0.901	0.951	0.638	0.601	0.655	0.400
0.25	0.05	0.270	0.229	0.274	0.005	0.019	0.025	0.004
0.25	0.10	0.625	0.576	0.620	0.109	0.093	0.110	0.084
0.25	0.15	0.895	0.841	0.853	0.349	0.307	0.367	0.288
0.25	0.20	0.955	0.909	0.967	0.628	0.580	0.617	0.420
0.50	0.05	0.234	0.218	0.240	0.009	0.018	0.018	0.007
0.50	0.10	0.596	0.542	0.616	0.129	0.099	0.138	0.111
0.50	0.15	0.855	0.813	0.867	0.380	0.347	0.387	0.291
0.50	0.20	0.911	0.929	0.923	0.624	0.584	0.631	0.412
0.75	0.05	0.258	0.227	0.264	0.005	0.018	0.024	0.007
0.75	0.10	0.612	0.562	0.621	0.147	0.128	0.152	0.124
0.75	0.15	0.870	0.851	0.872	0.383	0.340	0.380	0.301
0.75	0.20	0.928	0.914	0.930	0.640	0.612	0.641	0.462
$n_E = n_R = n_P = 100$								
0.0	0.05	0.547	0.543	0.532	0.105	0.102	0.101	0.004
0.0	0.10	0.896	0.901	0.903	0.531	0.508	0.545	0.090
0.0	0.15	0.988	0.987	0.986	0.900	0.887	0.891	0.480
0.0	0.20	0.999	0.999	0.999	0.989	0.991	0.991	0.851
0.25	0.05	0.521	0.493	0.525	0.097	0.098	0.097	0.003
0.25	0.10	0.888	0.880	0.881	0.475	0.469	0.464	0.086
0.25	0.15	0.984	0.984	0.983	0.858	0.869	0.874	0.463
0.25	0.20	1.000	1.000	0.999	0.991	0.989	0.986	0.858
0.50	0.05	0.494	0.490	0.493	0.087	0.102	0.085	0.004
0.50	0.10	0.877	0.871	0.875	0.453	0.485	0.466	0.081
0.50	0.15	0.985	0.988	0.983	0.885	0.881	0.873	0.467
0.50	0.20	1.000	1.000	1.000	0.991	0.993	0.995	0.866
0.75	0.05	0.508	0.498	0.502	0.088	0.106	0.084	0.003
0.75	0.10	0.889	0.908	0.906	0.513	0.557	0.529	0.086
0.75	0.15	0.997	0.995	0.997	0.937	0.934	0.936	0.468

λ	ξ	CBM	CPA	CBD	UCBM	UCPA	UCBD	Frequentist
0.75	0.20	1.000	1.000	1.000	0.997	0.998	1.000	0.852
$n_E = n_R = n_P = 300$								
0.0	0.05	0.756	0.777	0.767	0.302	0.288	0.294	0.045
0.0	0.10	0.995	0.996	0.994	0.942	0.941	0.939	0.670
0.0	0.15	1.000	1.000	1.000	0.991	1.000	0.999	0.981
0.0	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.05	0.740	0.735	0.742	0.277	0.265	0.273	0.046
0.25	0.10	0.999	0.991	0.994	0.918	0.920	0.921	0.646
0.25	0.15	1.000	1.000	0.995	0.987	1.000	0.990	0.981
0.25	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	0.05	0.727	0.719	0.732	0.251	0.254	0.258	0.035
0.50	0.10	0.996	0.993	0.993	0.930	0.924	0.921	0.660
0.50	0.15	1.000	1.000	1.000	0.999	0.999	0.999	0.987
0.50	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.05	0.771	0.760	0.775	0.287	0.294	0.297	0.040
0.75	0.10	0.998	0.995	0.997	0.951	0.956	0.955	0.716
0.75	0.15	1.000	1.000	1.000	1.000	1.000	1.000	0.994
0.75	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 7:

Remission and Response as outcome in the depression trial of Higuchi et al. [35].

Outcome	Duloxetine	Paroxetine	Placebo
Remission	50	49	32
Response	80	78	56
Total	$n_E = 147$	$n_R = 148$	$n_P = 145$

Table 8:

Bayesian and frequentist decision in the Depression trial with Remission as outcome, where “0” stands for rejection and “1” stands for acceptance of NI (with AS). These values are obtained using $n_{p^H} = 145$,

$$n_{R^H} = 148, \pi_{p^H} = 0.221, \pi_{R^H} = 0.331.$$

λ	CBM Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	CBD Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	CPA Post. Prob.	δ_{AS}^F	δ_{NI}^F	CBM Decision	CBD Decision	CPA Decision	Frequentist Decision
0.00	0.791, 0.098, 0.098	0.845, 0.101, 0.101	0.787	0.097	0.097	1	1	1	0
0.25	0.718, 0.098, 0.073	0.843, 0.101, 0.076	0.717	0.097	0.073	0	1	0	0
0.50	0.591, 0.098, 0.049	0.832, 0.101, 0.051	0.592	0.097	0.049	0	1	0	0
0.75	0.425, 0.098, 0.025	0.800, 0.101, 0.025	0.426	0.097	0.024	0	1	0	0
1.00	0.255, 0.098, 0	0.773, 0.101, 0	0.256	0.097	0	0	1	0	0

Note that non-inferiority is established only when one jointly rejects both NI and AS null hypotheses, which is defined as the final decision rule. For CPA the NI and AS margins are not reported separately as it is identical to the CBM approach.

Table 9:

Posterior Estimates and 95% credible interval under CBD for Remission.

Parameter	Estimate	Credible Interval
π_P	0.225	(0.179, 0.274)
π_R	0.316	(0.263, 0.372)
π_E	0.382	(0.308, 0.460)
$\pi_E - \pi_R$	0.065	(-0.027, 0.158)
$\pi_R - \pi_P$	0.143	(0.070, 0.203)

Table 10:

Bayesian and frequentist decision in the Depression trial with Response as outcome, where “0” stands for rejection and “1” stands for acceptance of NI (with AS). These values are obtained using $n_{P^H} = 145$,

$$n_{R^H} = 148, \pi_{P^H} = 0.386, \pi_{R^H} = 0.527.$$

λ	CBM Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	CBD Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	CPA Post. Prob.	δ_{AS}^F	δ_{NI}^F	CBM Decision	CBD Decision	CPA Decision	Frequentist Decision
0.00	0.896, 0.095, 0.095	0.906, 0.058, 0.058	0.895	0.100	0.100	1	1	1	0
0.25	0.842, 0.095, 0.072	0.903, 0.058, 0.043	0.843	0.100	0.075	1	1	1	0
0.50	0.785, 0.095, 0.048	0.905, 0.058, 0.029	0.775	0.100	0.050	1	1	1	0
0.75	0.720, 0.095, 0.024	0.907, 0.058, 0.014	0.689	0.100	0.025	0	1	0	0
1.00	0.693, 0.095, 0	0.899, 0.058, 0	0.590	0.100	0	0	1	0	0

Note that non-inferiority is established only when one jointly rejects both NI and AS null hypotheses, which is defined as the final decision rule. For CPA the NI and AS margins are not reported separately as it is identical to the CBM approach.

Table 11:

Posterior Estimates and 95% credible interval under CBD for Response.

Parameter	Estimate	Credible Interval
π_P	0.370	(0.315, 0.425)
π_R	0.484	(0.425,0.545)
π_E	0.624	(0.546,0.699)
$\pi_E - \pi_R$	0.139	(0.038, 0.239)
$\pi_R - \pi_P$	0.114	(0.030, 0.196)