Published in final edited form as:

Stat Biopharm Res. 2019 ; 11(1): 34–43. doi:10.1080/19466315.2018.1554504.

Bayesian Approach for Assessing Non-inferiority in Three-arm Trials for Risk Ratio and Odds Ratio

Shrabanti Chowdhury^{1,2}, Ram C. Tiwari³, Samiran Ghosh^{1,2,**}

¹Department of Family Medicine & Public Health Sciences, Wayne State University

²Center of Molecular Medicine and Genetics, Wayne State University

³Division of Biostatistics, CDRH, FDA, USA

Abstract

In this paper we consider three-arm non-inferiority (NI) trial that includes an experimental, a reference, and a placebo arm. While for binary outcomes the risk difference (RD) is the most common and well explored functional form for testing efficacy (or effectiveness), recent FDA guideline suggested other measures such as relative risk (RR) and odds ratio (OR) on the basis of which NI of an experimental treatment can be claimed. However, developing test based on these different functions of binary outcomes are challenging since the construction and interpretation of NI margin for such functions are not trivial extensions of RD based approach. Recently, we have proposed Frequentist approaches for testing NI for these functionals. In this article we further develop Bayesian approaches for testing NI based on effect retention approach for RR and OR. Bayesian paradigm provides a natural path to integrate historical trials' information, as well as it allows the usage of patients'/clinicians' opinions as prior information via sequential learning. In addition we discuss, in detail, the sample size/power calculation which could be readily used while designing such trials in practice.

Keywords

Assay Sensitivity; Binary Outcome; Fraction Margin; Markov chain Monte Carlo; Non-inferiority Margin; Risk/Odds Ratio; Three-arm Trial

1 Introduction

In the presence of established treatments/therapies, Active-controlled Non-inferiority (NI) trial designs are attractive ethical alternatives to placebo-controlled Randomized Control Trials (RCTs). A slightly less efficacious treatment may be preferable to a group of patients in view of lower toxicity, less intensive side effects, ease of delivery and other less debilitating factors. NI trials are intended to show if the new intervention retains a substantial portion of the active control effect, dictated by a pre-specified margin, often termed as NI margin (δ), yet possesses other attractive properties compared to established treatment regime. Such NI margin must be prospectively defined and should be so chosen to

^{**}Corresponding author: sghos@med.wayne.edu.

reflect maximum acceptable extent of clinical non-inferiority of an experimental treatment. Further detailed discussion on the construction and desirable properties of NI margin can be found in the regulatory guidelines (FDA (2016), ICHE9 (2009), ICHE10 (2009)) and references (e.g., Althunian et al. (2017), Schumi and Wittes (2011), Brown et al. (2006), Hung and Wang (2004)). NI trials may or may not include a placebo arm due to ethical reasons. Such two-arm placebo-free NI trials make two important assumptions regarding Assay Sensitivity (ICHE9 (2009), ICHE10 (2009)) and Constancy, and depend heavily on external validations (D'Agostino et al. (2003) and FDA (2016)) and several other limiting factors as specified in Kieser and Stucke (2016). Assay sensitivity (AS) refers to the ability of the trial to distinguish between effective and ineffective treatments. To alleviate some of these issues, it is recommended by EMA (2005) to include a placebo arm in the current trial, when ethically acceptable as well as practically feasible. This results in a three-arm “gold-standard” design that has greater confidence concerning AS and lesser concern on external validity.

For three-arm trial in the Frequentist setup, fraction margin approach (Pigeot et al., 2003) is popularly used for NI testing, where NI margin is adaptively constructed as the pre-specified negative fraction of the unknown effect size of the reference treatment over placebo in the current three-arm trial. This approach was extended by Kieser and Friede (2007) for the binary outcomes considering risk difference (RD) as the chosen function of interest, while Chowdhury et al. (2017) extended it for the same function using a Bayesian approach. In a very recent paper Ghosh et al. (2018) proposed a Bayesian approach for RD albeit under the joint testing of AS and NI. For binary outcomes, although RD is the simplest functional form, as mentioned in the recent FDA guidance (FDA (2016), Page 24), there are other functionals (e.g., risk ratio (RR), odds ratio (OR), etc.) which could be used for assessing treatment effect (Hashemi et al., 1997) and hence in the context of NI as well. Under the NI setup, there exists published work for odds ratio for two-arm trial under the Frequentist's approach, see for example Hilton (2010) and Rousson and Seifert (2008), however, to the best of our knowledge, no work exists for RR/OR for three-arm trial. Recently, Chowdhury et al. (2018b) proposed a novel Frequentist's approach incorporating the condition of AS for developing three-arm NI tests for RR and OR. Also very recently Chowdhury et al. (2018a) developed a Bayesian approach of testing NI for RR/OR in two arm trial. Since NI trials involve treatments that have been well studied in the past, the availability and usage of historical information in the current NI trial is an advantage (Ghosh et al., 2011; Gamalo et al., 2011; Ghosh et al., 2018). Bayesian paradigm provides a natural path to incorporate the useful information from various historical trials and then to combine them with the current trial, thus possibly reducing sample size and other cost burden. In this paper, we propose a fully Bayesian and two approximation-based Bayesian approaches for both RR and OR. We also provide the sample size calculation under Bayesian methods to achieve a desired level of power.

The rest of the article is organized as follows. In Section 2, we give the NI hypothesis for RR and OR for three-arm NI trial and discuss briefly the Frequentist's methods for testing. In Section 3, we propose three Bayesian methods, incorporating the condition of assay sensitivity, to design and perform the analysis of NI trial for both the functionals. In Section 4, we discuss the power and sample size calculation and details of simulation studies. Finally

in Section 5, we apply our proposed methods on a published clinical trial dataset. We conclude the article with discussions in Section 6. All proofs are provided in a separate supplementary file.

2 Frequentist Approach for NI Testing for Different Functionals

For a three-arm trial, fraction margin approach (Pigeot et al., 2003; Kieser and Friede, 2007) is popularly used for testing NI hypothesis and finding the corresponding decision rule. We begin our illustration borrowing the notations from Kieser and Friede (2007). Denote the primary end-points from the Placebo (P), Reference (R) and the Experimental (E) treatment in the current trial by X_P , X_R , and X_E respectively, each following $Bin(n_l, \pi_l)$, where π_l is the probability of success and n_l is the sample size for the l^{th} arm, $l \in \{P, R, E\}$. Without loss of generality, we assume that higher response probabilities indicate greater treatment benefits. Gamalo et al. (2011) used the two-arm fixed margin approach for NI testing considering RD as the functional of interest. Pigeot et al. (2003) formulated the three-arm NI hypothesis for continuous outcome, whereas Kieser and Friede (2007) formulated the same for binary outcomes under fraction margin approach. In this set up, the construction of $\delta (< 0)$ can be mathematically expressed as $\delta = f(\pi_R - \pi_P)$, where f is a negative fraction, $f \in [-1, 0]$, assuming the condition of assay sensitivity, that is, $(\pi_R - \pi_P > 0)$ holds. We reformulate the NI hypothesis for RR and OR using a general functional form as

$$H_0: \psi(\pi_E, \pi_R) \leq \delta \text{ vs. } H_1: \psi(\pi_E, \pi_R) > \delta, \quad (2.1)$$

where $\delta \in [0, 1]$ denotes a pre-chosen portion of the unknown effect size of the activecontrol over placebo and $\psi(\cdot)$ is an appropriate function of choice (e.g., RR, OR). To formulate the NI hypothesis for RR and OR, we follow Wangge et al. (2013), where the formula for calculating δ (which is same as the quantity M_2 in their paper) is given as $M_2 = (1/M_1)^{1-\text{preserved effects}}$, where M_1 is the effect of the active control relative to the placebo that is still present in the current trial. Along the same line, for the RR (or OR) scale we propose the construction of δ as $\delta = (1/\psi(\pi_R, \pi_P))^{-f} = (1/\psi(\pi_R, \pi_P))^{1-\theta}$, where $f \in [-1, 0]$ is the loss of effect of the experimental drug compared to active control effect, and $\theta = 1 + f$ is the preservation effect; that is, the proportion of the active control effect retained by the

experimental drug in the current trial. Now for the specific case of RR, $\psi(\pi_E, \pi_R) = \frac{\pi_E}{\pi_R}$ and

$$\psi(\pi_R, \pi_P) = \frac{\pi_R}{\pi_P}, \text{ while for OR, } \psi(\pi_E, \pi_R) = \frac{\pi_E/(1-\pi_E)}{\pi_R/(1-\pi_R)} \text{ and } \psi(\pi_R, \pi_P) = \frac{\pi_R/(1-\pi_R)}{\pi_P/(1-\pi_P)}. \text{ Note that}$$

for both RR and OR, the NI margin satisfies two important boundary conditions: for $f = -1$, the active control loses its practical significance over placebo, hence the test reduces to the simple superiority test of π_E over π_P while for $f = 0$, the test becomes superiority test of π_E over π_R . Below, we give the NI hypothesis for RR and OR respectively:

$$\text{RR: } H_0: \frac{\pi_E}{\pi_R} \leq \left(\frac{\pi_P}{\pi_R}\right)^{(1-\theta)} \text{ vs. } H_1: \frac{\pi_E}{\pi_R} > \left(\frac{\pi_P}{\pi_R}\right)^{(1-\theta)}. \quad (2.2)$$

$$\text{OR: } H_0: \frac{\pi_E'(1-\pi_E)}{\pi_R'(1-\pi_R)} \leq \left(\frac{\pi_P'(1-\pi_P)}{\pi_R'(1-\pi_R)} \right)^{(1-\theta)} \text{ vs. } H_1: \frac{\pi_E'(1-\pi_E)}{\pi_R'(1-\pi_R)} > \left(\frac{\pi_P'(1-\pi_P)}{\pi_R'(1-\pi_R)} \right)^{(1-\theta)} \quad (2.3)$$

Since logarithm transformations make the data conform more closely to the Normal distribution giving better asymptotic performance, we also give the NI hypothesis for RR and OR taking logarithm of both sides of (2.2) and (2.3) respectively:

$$\text{RR: } H_0: \log(\pi_E) - \theta \log(\pi_R) - (1-\theta) \log(\pi_P) \leq 0 \text{ vs. } H_1: \log(\pi_E) - \theta \log(\pi_R) - (1-\theta) \log(\pi_P) > 0; \quad (2.4)$$

$$\begin{aligned} \text{OR: } H_0: \log\left(\frac{\pi_E}{(1-\pi_E)}\right) - \theta \log\left(\frac{\pi_R}{(1-\pi_R)}\right) - (1-\theta) \log\left(\frac{\pi_P}{(1-\pi_P)}\right) &\leq 0 \text{ vs. } \\ H_1: \log\left(\frac{\pi_E}{(1-\pi_E)}\right) - \theta \log\left(\frac{\pi_R}{(1-\pi_R)}\right) - (1-\theta) \log\left(\frac{\pi_P}{(1-\pi_P)}\right) &> 0. \end{aligned} \quad (2.5)$$

Next, we briefly discuss the Frequentist approaches for NI testing for RR and OR that have already been proposed in Chowdhury et al. (2018b). The NI hypothesis in (2.4) and (2.5) can be written in a general form as

$$H_0: g(\pi_E) - \theta g(\pi_R) - (1-\theta)g(\pi_P) \leq 0 \text{ vs. } H_1: g(\pi_E) - \theta g(\pi_R) - (1-\theta)g(\pi_P) > 0. \quad (2.6)$$

Thus for RR: $g(\pi_I) = \log(\pi_I)$, and for OR: $g(\pi_I) = \log(\pi_I/(1-\pi_I))$, $I \in \{E, R, P\}$.

Chowdhury et al. (2018b) proposed the marginal Frequentist's approach of NI testing for RR and OR following the existing guideline developed by Pigeot et al. (2003) and Kieser and Friede (2007). The test is based on the assumption of asymptotic normality of the Frequentist test statistic $T = g(\hat{\pi}_E) - \theta g(\hat{\pi}_R) - (1-\theta)g(\hat{\pi}_P)$, $\hat{\pi}_I$ being the MLE of π_I , $I \in \{E, R, P\}$. However, since the marginal approach does not take into account the AS condition, it leads to biased estimates of power and/or sample size. Chowdhury et al. (2018b) also proposed the conditional approach of NI testing, considering the test statistic based on the restricted maximum likelihood estimate (RMLE), thus incorporating the AS condition in the Frequentist's statistic itself. This leads to the modified test statistic for NI testing:

$W = g(\hat{\pi}_E) - \theta g(\hat{\pi}_R) - (1-\theta)g(\hat{\pi}_P) | (g(\hat{\pi}_R) - g(\hat{\pi}_P) > 0)$. Under the asymptotic normality (AN)

of W , $\frac{W - \mu_W}{\sigma_W} \sim AN(0, 1)$, where $E[W] = \mu_W$ and $V[W] = \sigma_W^2$. The test is rejected if $W > k^*$,

where k^* is obtained by assuming a test of size α : $P_{H_0}(W > k^*) = \alpha$. We express the sample

size in the reference (n_R) and the experimental (n_E) arms as the ratio r_1 and r_2 respectively of the sample size $n_P = n$, say, in the placebo arm such that $n_P : n_R : n_E = 1 : n_R/n_P : n_E/n_P = 1 : r_1 : r_2$. Here, r_1 and r_2 are known positive quantities that determine the allocation ratio of the sample sizes in the arms R and E , respectively, relative to the arm P . Denoting n_P by n , the total sample size, thus, would be $N = n(1 + r_1 + r_2)$. The sample size $n_P = n$ (for the arm P) is calculated from the following equation to achieve a power of at least $100(1 - \beta)\%$

$$(\text{Chowdhury et al., 2018b}): P_{H_1}(W > k^*) \geq 1 - \beta \Rightarrow \Phi\left(\frac{k^* - \mu_w^{alt}}{\sigma_w^{alt}}\right) \leq \beta, \text{ where } \mu_w^{alt} \text{ and } (\sigma_w^{alt})^2 \text{ are}$$

the mean and variance of W under the alternative hypothesis in (2.6).

3 Proposed Bayesian Approaches for Non-Inferiority

In an NI trial with three-arms, if the condition of AS is satisfied, it assures that the NI study is being carried out under more or less the same circumstance as that of the former studies, in which efficacy of reference drug was tested. Gamalo et al. (2011) developed Bayesian procedures for NI testing in a two-arm trial for RD that allows the incorporation of the historical data on the active control via the use of informative priors. Chowdhury et al. (2017) developed NI testing in a three-arm trial for RD under Bayesian paradigm considering fully Bayesian and approximation-based Bayesian approaches, incorporating the AS condition. In this Section, we propose a fully Bayesian and two approximation-based Bayesian test procedures. However, we emphasize the fact that, for achieving better accuracy/estimate of the sample size one should consider fully Bayesian approach almost always, although it comes at a cost of extensive computation. Note that, parallel to the conditional Frequentist approach as discussed in Section 2, we incorporate the AS condition ($\pi_R - \pi_P > 0$) in our Bayesian approaches as well.

3.1 Fully Bayesian Approach

We consider the conjugate prior, where the AS condition ($\pi_R - \pi_P > 0$) or equivalently $g(\pi_R) - g(\pi_P) > 0$ is incorporated explicitly parallel to the proposed conditional Frequentist's approach in Chowdhury et al. (2018b). In the conjugate prior setting, we assume a Beta distribution as prior to the proportion of successes in each arm of the trial, that is, $\pi_I \sim \text{Beta}(\alpha_I, \beta_I)$, $I \in \{E, R, P\}$, where α_I, β_I are fixed hyper-parameters which can take any value on the positive real line $\mathbb{R}^+$. After incorporating the assumption of AS, the joint prior distribution in the three-arms becomes:

$$f(\pi_E, \pi_R, \pi_P) \propto I(\pi_R > \pi_P) \prod_{I \in \{E, R, P\}} f(\pi_I | \alpha_I, \beta_I),$$

where $f(\pi_I | \alpha_I, \beta_I)$ is the density of the standard Beta distribution given by

$f(\pi_I | \alpha_I, \beta_I) \propto \pi_I^{\alpha_I - 1} (1 - \pi_I)^{\beta_I - 1}$, $0 < \pi_I < 1$. Under the well known Beta-Binomial conjugacy, posterior distribution of π_I , $I \in \{E, R, P\}$, is also a Beta distribution. Their joint posterior distribution is given by,

$$f(\pi_E, \pi_R, \pi_P | X_E, X_R, X_P, \alpha_I, \beta_I) \propto I(\pi_R > \pi_P) \prod_{I \in \{E, R, P\}} \pi_I^{\alpha_I + x_I - 1} (1 - \pi_I)^{\beta_I + n_I - x_I - 1},$$

$0 < \pi_I < 1$. The Markov Chain Monte Carlo (MCMC) samplers can be easily generated from the joint posterior, satisfying $\pi_R > \pi_P$. The choice of hyper-parameters ($\alpha_I, \beta_I, I \in \{E, R, P\}$) are driven by the precision of available prior information. For example, in the absence of prior information, a flat prior may be assigned by choosing hyper-parameter values resulting in large variance. The mean (μ), mode (μ^0) and variance (σ^2) of $Beta(\alpha, \beta)$ are given as, $\mu = \frac{\alpha}{\alpha + \beta}$, $\mu^0 = \frac{\alpha - 1}{\alpha + \beta - 2}$, $\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$. The informative prior can be

obtained by equating the mean (or mode) with the parameter $\pi_I, I \in \{E, R, P\}$ and making the variance rather smaller.

In determining NI of the experimental drug relative to the active control, the investigator picks a value of $\theta \in [0.5, 1)$ that represents clinically acceptable effect size relative to the placebo. The Bayesian decision rule to claim that the experimental drug is non-inferior to the reference is given by

$$P\left(H_1: \frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} > \theta | g(\pi_R) - g(\pi_P) > 0, X\right) > p^*, \quad (3.1)$$

where p^* is some clinically meaningful pre-specified cut-off, $X = \{X_E, X_R, X_P\}$ denotes the relevant data. The above probability can be calculated empirically by generating samplers from the posterior distribution of $(\pi_I | X)$, and then calculating the functional $g(\cdot), I \in \{E, R, P\}$, where the forms of $g(\cdot)$ for RR and OR are defined earlier. The estimated probability is given by

$$\hat{P}\left(\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} > \theta | g(\pi_R) > g(\pi_P), X\right) \approx \frac{1}{T} \sum_{t=1}^T I\left(\frac{g(\pi_E^t) - g(\pi_P^t)}{g(\pi_R^t) - g(\pi_P^t)} > \theta | g(\pi_R^t) > g(\pi_P^t), X\right), \quad (3.2)$$

where π_E^t, π_R^t and π_P^t denote the t^{th} MCMC sample drawn from the appropriate posterior distribution satisfying $(g(\pi_R^t) - g(\pi_P^t) > 0)$, with sufficiently large T , and $g(\pi_I^t)$ is the function $g(\cdot)$ evaluated at the sample value $\pi_I^t, I \in \{E, R, P\}, t = 1, \dots, T$.

3.2 Normal Approximation to the Prior

We propose an approximation-based Bayesian approach for NI testing incorporating the AS condition, that gives closed form of the posterior probability and hence saves the computation time for the MCMC sample generation from the posterior distribution. In this

Section, we derive the test procedure and sample size calculation formula for both RR and OR.

Under the Binomial responses, the marginal Frequentist's test statistic

$T(X) = g(\hat{\pi}_E) - \theta g(\hat{\pi}_R) - (1 - \theta)g(\hat{\pi}_P)$ is used for testing the hypothesis expressed in the

general form in (2.6). Under asymptotic normality, we have $T(X)|\mu_T \sim AN(\mu_T, \sigma_T^2)$, where μ_T

$= g(\pi_E) - \theta g(\pi_R) - (1 - \theta)g(\pi_P)$ and $\sigma_T^2 = \sigma_E^2 + \theta^2 \sigma_R^2 + (1 - \theta)^2 \sigma_P^2$, with $\sigma_l^2 = \frac{1 - \pi_l}{n_l \pi_l}$ for RR,

and $\sigma_l^2 = \frac{1}{n_l \pi_l (1 - \pi_l)}$ for OR; $l \in \{E, R, P\}$. Under the conjugate Beta prior, de-note the mean

and variance of π_l by μ_l^0 and $(\sigma_l^0)^2$, $l \in \{E, R, P\}$. The mean $\mu_l^g = E(g(\pi_l))$ can be

approximated up to second order term of the Taylor series expansion;

$\mu_l^g = E(g(\pi_l)) \approx g(\mu_l^0) + \frac{(\sigma_l^0)^2}{2} g''(\mu_l^0)$ and the variance for $g(\pi_l)$ can be approximated by the

delta-method as $(\sigma_l^g)^2 = \text{Var}(g(\pi_l)) = \text{Var}(\pi_l) g'^2(\pi_l)$ at $\pi_l = E(\pi_l)$, $l \in \{E, R, P\}$. For RR,

$\mu_l^g = \log(\mu_l^0) - \frac{(\sigma_l^0)^2}{2(\mu_l^0)^2}$ and $(\sigma_l^g)^2 = \frac{(\sigma_l^0)^2}{(\mu_l^0)^2}$ while for OR, $\mu_l^g = \log\left(\frac{\mu_l^0}{1 - \mu_l^0}\right) + \frac{(\sigma_l^0)^2}{2} \left(\frac{1}{(1 - \mu_l^0)^2} - \frac{1}{(\mu_l^0)^2}\right)$

and $(\sigma_l^g)^2 = \frac{(\sigma_l^0)^2}{(\mu_l^0)^2 (1 - \mu_l^0)^2}$. Putting Normal prior on μ_T we have $\mu_T \sim AN(\mu^*, \sigma^{*2})$, where

$\mu^* = E(\mu_T) = \mu_E^g - \theta \mu_R^g - (1 - \theta) \mu_P^g$, and $\sigma^{*2} = (\sigma_E^g)^2 + \theta^2 (\sigma_R^g)^2 + (1 - \theta)^2 (\sigma_P^g)^2$. Next we bring

in the condition of assay sensitivity, $(g(\pi_R) - g(\pi_P) > 0)$. So instead of taking prior on μ_T ,

we take prior on $\nu_T \equiv \mu_T | (g(\pi_R) - g(\pi_P) > 0)$. Assume that $\nu_T \sim AN(\mu_v^*, \sigma_v^{*2})$. Then, the

posterior distribution of ν_T becomes $\nu_T | T(X) \sim AN(\tilde{\mu}_T, \tilde{\sigma}_T^2)$. We refer to Arnold and Beaver

(1993) for the detailed derivation of μ_v^*, σ_v^{*2} .

Lemma 3.2.1 Under conditional normal approximation, the mean μ_v^* and variance σ_v^{*2} of ν_T

$= g(\pi_E) - \theta g(\pi_R) - (1 - \theta)g(\pi_P) | (g(\pi_R) - g(\pi_P) > 0)$ are given by

$$\begin{aligned} \mu_v^* &= \mu_{\eta EP} + \sigma_{\eta EP} \frac{\rho}{c} \phi(d) - \theta \left(\mu_{\eta RP} + \sigma_{\eta RP} \frac{1}{c} \phi(d) \right) \\ \sigma_v^{*2} &= \sigma_{\eta EP}^2 \left[1 + \frac{\rho^2}{c} a \phi(d) - \left(\frac{\rho}{c} \phi(d) \right)^2 \right] + \theta^2 \sigma_{\eta RP}^2 \left[1 - \frac{\phi(d)}{c} \left(\frac{\phi(d)}{c} - d \right) \right] \\ &\quad - 2\theta \left[\sigma_{\eta EP} \sigma_{\eta RP} \frac{\rho}{c} (c + a \phi(d)) + \sigma_{\eta EP} \mu_{\eta RP} \frac{\rho}{c} \phi(d) + \sigma_{\eta RP} \mu_{\eta EP} \frac{1}{c} \phi(d) \right. \\ &\quad \left. + \mu_{\eta EP} \mu_{\eta RP} - \left(\mu_{\eta EP} + \sigma_{\eta EP} \frac{\rho}{c} \phi(d) \right) \left(\mu_{\eta RP} + \sigma_{\eta RP} \frac{1}{c} \phi(d) \right) \right], \end{aligned} \quad (3.3)$$

where $\mu_{\eta RP} = \mu_R^g - \mu_P^g$, $\mu_{\eta EP} = \mu_E^g - \mu_P^g$, $\sigma_{\eta EP}^2 = (\sigma_E^g)^2 + (\sigma_P^g)^2$, $\sigma_{\eta RP}^2 = (\sigma_R^g)^2 + (\sigma_P^g)^2$,
 $\rho = \frac{(\sigma_P^g)^2}{\sigma_{\eta EP} \sigma_{\eta RP}}$, $d = -\frac{\mu_{\eta RP}}{\sigma_{\eta RP}}$ and $c = 1 - \Phi(d)$, μ_l^g and $(\sigma_l^g)^2$ are functions of the mean and
 variance of Beta (α_l, β_l) , $l \in \{E, R, P\}$.

Proof: See A.1 in the supplementary material.

The Bayesian decision rule for claiming the experimental treatment to be non-inferior is parallel to Gamalo et al. (2014) and is given by: $P(v_T = 0 | X_E, X_R, X_P) > p^*$, where p^* is a pre-specified constant usually chosen to be 0.975 or 0.95.

3.3 Normal Approximation to the Posterior

As in the previous approximation approach, we consider similar setting: $X_l \sim \text{Bin}(n_l, \pi_l)$ and $\pi_l \sim \text{Beta}(\alpha_l, \beta_l)$, $l \in \{E, R, P\}$. Instead of approximating the Frequentist's statistic T and the prior by Normal distribution, approximation can be also made to the posterior distribution of $g(\pi_l) | X_l$, $l \in \{E, R, P\}$. Our posterior approximation approach follows the lines of the posterior normal approximation proposed by Gamalo et al. (2011) for binary responses, albeit for two arm NI testing. Bringing in the condition of assay sensitivity, we derive the posterior mean and variance of

$v_T^P \equiv (g(\pi_E) - g(\pi_P)) - \theta(g(\pi_R) - g(\pi_P)) | (g(\pi_R) - g(\pi_P) > 0, X)$. The posterior mean and variance

of $\pi_l | X_l$ are respectively given by: $\mu_l^P = \frac{\alpha_l + X_l}{\alpha_l + \beta_l + n_l}$ and $(\sigma_l^P)^2 = \frac{(\alpha_l + X_l)(\beta_l + n_l - X_l)}{(\alpha_l + \beta_l + n_l)^2(\alpha_l + \beta_l + n_l + 1)}$.

Hence, the posterior mean of $g(\pi_l) | X_l$ can be approximated similar to that has been done in

the prior approximation approach as $\mu_l^{gP} = E(g(\pi_l) | X_l) \approx g(\mu_l^P) + \frac{(\sigma_l^P)^2}{2} g''(\mu_l^P)$ and the posterior variance as $(\sigma_l^{gP})^2 = \text{Var}(g(\pi_l) | X_l) = \text{Var}(\pi_l | X_l) g'^2(\pi_l)$ at $\pi_l = E(\pi_l | X_l)$. Specifically for RR,

$$\mu_l^{gP} = \log(\mu_l^P) - \frac{(\sigma_l^P)^2}{2(\mu_l^P)^2} \text{ and } (\sigma_l^{gP})^2 = \frac{(\sigma_l^P)^2}{(\mu_l^P)^2} \text{ while for OR,}$$

$$\mu_l^{gP} = \log\left(\frac{\mu_l^P}{1 - \mu_l^P}\right) + \frac{(\sigma_l^P)^2}{2} \left(\frac{1}{(1 - \mu_l^P)^2} - \frac{1}{(\mu_l^P)^2} \right) \text{ and } (\sigma_l^{gP})^2 = \frac{(\sigma_l^P)^2}{(\mu_l^P)^2 (1 - \mu_l^P)^2}. \text{ Thus, under}$$

asymptotic normality we have $v_T^P \sim \text{AN}(\mu_v^P, (\sigma_v^P)^2)$. We again refer to Arnold and Beaver (1993) for the derivation of μ_v^P and $(\sigma_v^P)^2$.

Lemma 3.3.1 Under conditional normal approximation, the mean μ_v^P and variance $(\sigma_v^P)^2$ of $v_T^P = (g(\pi_E) - g(\pi_P)) - \theta(g(\pi_R) - g(\pi_P)) | (g(\pi_R) - g(\pi_P) > 0, X)$ are given by

$$\begin{aligned}
 \mu_v^P &= \mu_{\eta EP} + \sigma_{\eta EP} \frac{\rho}{c} \phi(d) - \theta \left(\mu_{\eta RP} + \sigma_{\eta RP} \frac{1}{c} \phi(d) \right) \\
 (\sigma_v^P)^2 &= \sigma_{\eta EP}^2 \left[1 + \frac{\rho^2}{c} a \phi(d) - \left(\frac{\rho}{c} \phi(d) \right)^2 \right] + \theta^2 \sigma_{\eta RP}^2 \left[1 - \frac{\phi(d)}{c} \left(\frac{\phi(d)}{c} - d \right) \right] \\
 &\quad - 2\theta \left[\sigma_{\eta EP} \sigma_{\eta RP} \frac{\rho}{c} (c + a \phi(d)) + \sigma_{\eta EP} \mu_{\eta RP} \frac{\rho}{c} \phi(d) + \sigma_{\eta RP} \mu_{\eta EP} \frac{1}{c} \phi(d) \right. \\
 &\quad \left. + \mu_{\eta EP} \mu_{\eta RP} - \left(\mu_{\eta EP} + \sigma_{\eta EP} \frac{\rho}{c} \phi(d) \right) \left(\mu_{\eta RP} + \sigma_{\eta RP} \frac{1}{c} \phi(d) \right) \right],
 \end{aligned} \tag{3.4}$$

where $\mu_{\eta RP} = \mu_R^{gP} - \mu_P^{gP}$, $\mu_{\eta EP} = \mu_E^{gP} - \mu_P^{gP}$, $\sigma_{\eta EP}^2 = (\sigma_E^{gP})^2 + (\sigma_P^{gP})^2$, $\sigma_{\eta RP}^2 = (\sigma_R^{gP})^2 + (\sigma_P^{gP})^2$, $\rho = \frac{(\sigma_P^{gP})^2}{\sigma_{\eta EP} \sigma_{\eta RP}}$, $d = -\frac{\mu_{\eta RP}}{\sigma_{\eta RP}}$ and $c = 1 - \Phi(d)$, μ_l^{gP} and $(\sigma_l^{gP})^2$ are functions of the mean and variances of $\pi_l | X_l \sim \text{Beta}(\alpha_l + X_l, \beta_l + n_l - X_l)$, $l \in \{E, R, P\}$.

Proof: See A.2 in the supplementary material.

The Bayesian decision rule to claim that the experimental treatment is non-inferior to the active control in the current trial is given by $P(v_T^P \geq 0) > p^* \Rightarrow \Phi\left(\frac{\mu_v^P}{\sigma_v^P}\right) > p^*$.

3.4 Sample Size under Proposed Approaches

3.4.1 Fully Bayesian—We obtain the power of the test by repeating the calculation of the estimated probability in (3.2), n^* times (usually 1000 or 5000), and then finding the proportion of times NI of the test drug is claimed out of n^* . The power function is obtained by varying the value of π_E for given values of π_R and π_P such that $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$ belongs to a suitable range under the alternative, H_1 . The ratio $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$ is so chosen that for NI testing under H_0 it equals $\theta \in [0.5, 1)$ and exceeds θ under H_1 . Denote π_E by π_E^{alt} under H_1 , and the estimated power of the test can be calculated as

$$\widehat{\text{Power}} = \frac{\# \text{ of times } P\left(g(\pi_E^{alt}) - \theta g(\pi_R) - (1 - \theta)g(\pi_P) > 0 \mid g(\pi_R) - g(\pi_P) > 0, X\right) > p^*}{n^*}.$$

Considering the allocation ratio as discussed in Section 2, the minimum sample size n of the arm P can be obtained by setting the power at $100(1 - \beta)\%$. The sample size in the other two arms can be obtained from the respective allocation ratios. We note that, since the power and sample size calculations involve generating samples from posterior distribution, there could be possibly some sampling fluctuation.

3.4.2 Normal Approximation to the Prior—The sample size n of the arm P under Bayesian prior approximation approach can be calculated by satisfying the following two conditions (see also Gamalo et al. (2014))

$$(C1): P[P(v_T \geq 0|X) \geq p^* | H_0] \leq \alpha, (C2): P[P(v_T \geq 0|X) \geq p^* | H_1] \geq 1 - \beta,$$

where the probability in (C1) is the estimated Bayesian version of type-I error while that in (C2) is the estimated power of the test, β being the type-II error. The sample size n is determined from (C2) by fixing β to achieve at least 100 $(1 - \beta)$ % power and simultaneously satisfying (C1). As in the Frequentist's approach, we choose $\alpha = 0.025$. We note that

$$\begin{aligned} P(v_T \geq 0 | X_E, X_R, X_P) &= P\left(\frac{v_T - \tilde{\sigma}_T^2 \tilde{\mu}_T}{\tilde{\sigma}_T} > \frac{-\tilde{\sigma}_T^2 \tilde{\mu}_T}{\tilde{\sigma}_T}\right) > p^* \\ &\Leftrightarrow -\tilde{\sigma}_T \tilde{\mu}_T < z_{1-p^*} \Leftrightarrow T > -z_{1-p^*} \left(\frac{1}{\sigma_T^2} + \frac{1}{\sigma_v^2}\right)^{1/2} \sigma_T^2 - \frac{\mu_v^*}{\sigma_v^2} \sigma_T^2, \end{aligned}$$

where z_{1-p^*} is the 100 $(1 - p^*)$ % of the $N(0, 1)$ distribution. Now the power function is obtained by varying π_E , keeping the other parameters π_R , π_P and θ fixed. Let us denote μ_T and σ_T^2 by μ_T^{null} and $(\sigma_T^{null})^2$, respectively under H_0 , and similarly under H_1 denote the respective quantities by μ_T^{alt} and $(\sigma_T^{alt})^2$. Thus, condition (C1) can be rewritten in terms of T as

$$\begin{aligned} &P_{H_0} \left[T > z_{1-p^*} \left(\frac{1}{(\sigma_T^{null})^2} + \frac{1}{\sigma_v^{*2}} \right)^{1/2} \left((\sigma_T^{null})^2 - \frac{\mu_v^*}{\sigma_v^{*2}} (\sigma_T^{null})^2 \right) \right] \leq \alpha \\ \Leftrightarrow &P_{H_0} \left[\frac{T - \mu_T^{null}}{\sigma_T^{null}} > \left(-z_{1-p^*} \left(\frac{1}{(\sigma_T^{null})^2} + \frac{1}{\sigma_v^{*2}} \right)^{1/2} \left((\sigma_T^{null})^2 - \frac{\mu_v^*}{\sigma_v^{*2}} (\sigma_T^{null})^2 - \mu_T^{null} \right) / \sigma_T^{null} \right) \right] \leq \alpha \\ \Leftrightarrow &\Phi \left(z_{1-p^*} \left(\frac{1}{(\sigma_T^{null})^2} + \frac{1}{\sigma_v^{*2}} \right)^{1/2} \left(\sigma_T^{null} + \frac{\mu_v^*}{\sigma_v^{*2}} \sigma_T^{null} + \frac{\mu_T^{null}}{\sigma_T^{null}} \right) \right) \leq \alpha \end{aligned} \quad (3.5)$$

Similarly, condition (C2) becomes

$$\Phi \left(z_{1-p^*} \left(\frac{1}{(\sigma_T^{alt})^2} + \frac{1}{\sigma_v^{*2}} \right)^{1/2} \left(\sigma_T^{alt} + \frac{\mu_v^*}{\sigma_v^{*2}} \sigma_T^{alt} + \frac{\mu_T^{alt}}{\sigma_T^{alt}} \right) \right) \geq 1 - \beta. \quad (3.6)$$

Now n can be solved from equation (3.6) by setting $\beta = 20\%$ and simultaneously satisfying (3.5). We vary π_E^{alt} (which is included in μ_T^{alt}) to get minimum sample size satisfying at least 80% power for each π_E^{alt} . The sample size for the arms R and P can be obtained considering the allocation ratios r_1 and r_2 as discussed earlier.

3.4.3 Normal Approximation to the Posterior—Under the Bayesian posterior

approximation approach, we note that $\Phi\left(\frac{\mu_v^P}{\sigma_v^P}\right)$ involves the data (X_E, X_R, X_P) since the mean and variance of v_T^P are functions of the posterior distribution of $\pi_I | X_I, I \in \{E, R, P\}$. The

power of the test can be calculated as the proportion of times $\Phi\left(\frac{\mu_v^P}{\sigma_v^P}\right)$ exceeds p^* , out of n^* .

The power function is obtained by varying the value of π_E for given values of the other parameters. The estimated power of the test is obtained at $\pi_E = \pi_E^{alt}$ satisfying

$g(\pi_E^{alt}) > g(\pi_R) + \theta(g(\pi_R) - g(\pi_P))$ under H_1 , and can be calculated as

$$\widehat{Power} = \frac{\# \text{ of times } \Phi\left(\frac{\mu_v^P}{\sigma_v^P}\right) > p^*}{n^*}. \text{ The sample size of the respective treatment arms can be}$$

obtained by setting the power to be at least $100(1 - \beta)\%$ and from the allocation ratios as discussed in the Section 2.

4 Simulation Studies and Sample Size Determination

In this Section, we first evaluate the performance of the Bayesian procedures and then focus on the sample size calculation for testing NI at a desired power. We consider both balanced and unbalanced allocation of the total sample size in the three arms.

4.1 Simulation Set-up

The following simulation steps are conducted to calculate the type-I error and power of the tests, under the conditional Frequentist, fully Bayesian as well as two Bayesian approximation approaches. Under the Bayesian approaches, we assume a non-informative prior for $\pi_I, I \in \{E, R, P\}$ for both RR and OR. Considering a randomized trial we generate the power curves assuming equal allocation to the three arms; that is, $n_E = n_R = n_P = n$. In the following we give the simulation steps for the fully Bayesian approach. The steps for the approximation-based approaches are similar.

S1: Specify $n_E = n_R = n_P = n$, $\pi_I, I \in \{E, R, P\}$ with $\pi_P < \pi_R$, and θ and vary π_E such that $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$ belongs to a suitable range, to generate X_E, X_R and X_P

S2: For a given value of $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$ or equivalently π_E , generate the data X_I from Binomial distribution $Bin(n_I, \pi_I), I \in \{E, R, P\}$.

- S3:** Generate T many MCMC samples from the posterior distribution given in Section 3 satisfying the condition $\pi_R > \pi_P$ or equivalently $g(\pi_R) > g(\pi_P)$.
- S4:** For the t^{th} posterior sample calculate the ratio $\frac{g(\pi_E^t) - g(\pi_P^t)}{g(\pi_R^t) - g(\pi_P^t)}$ and calculate the posterior probability: $P\left(\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} > \theta | X\right) \approx \frac{1}{T} \sum_{t=1}^T I\left(\frac{g(\pi_E^t) - g(\pi_P^t)}{g(\pi_R^t) - g(\pi_P^t)} > \theta | X\right)$.
- S5:** Bayesian criterion: Initialize COUNTS to 0; if $P\left(\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} > \theta | X\right) > p^*$, increase COUNTS by 1; otherwise 0.
- S6:** Go back to Step 2 and repeat the simulation n^* (a large number chosen apriori) times:
- Calculate type-I error by dividing COUNTS by n^* for π_E satisfying $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} = \theta$.
 - Calculate power by dividing COUNTS by n^* for π_E satisfying $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)} > \theta$.
- S7:** The power curve is generated for a sequence of values of π_E , such that $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$ belongs to a suitable wide enough range.

Note that for the conditional Frequentist and Bayesian approximation approaches Step 3 is not needed and Step 4 and Step 5 need to be replaced by the corresponding decision criterion as given in Section 2, 3.1, 3.2, and 3.3 respectively.

4.2 Sample Size in Simulation

Before generating the sample size tables we generate the power curves under fully Bayesian approach to get an idea about the operating characteristics of the proposed methods. For the approximation-based approaches the power curves can be generated similarly. For the conjugate Beta prior, since the posterior is available in closed form, we chose the number of posterior samplers T to be 1000. Throughout the simulation study, we consider the following

specification of the parameters: $\pi_R = 0.7$, $\pi_P = 0.1$ and we set π_E such that $\frac{g(\pi_E) - g(\pi_P)}{g(\pi_R) - g(\pi_P)}$

gives a wide enough range for exploration purpose for both RR and OR. We chose n^* to be 5000. We take the value of the common sample size $n_E = n_R = n_P = n$ to be 100. Unequal allocation is also possible and is presented in the sample size tables (Table 1 and Table 2). Another important criteria is the choice of p^* which we fix at 0.975. However, as reported in Gamalo et al. (2012) this choice could give too restrictive type-I error in Bayesian context

and Bayesian calibration can be performed to alleviate this; however, this is not pursued here.

In Figure 1, we plot the power curves corresponding to three different values of θ : 0.8, 0.7 and 0.6 under conjugate prior for each treatment arm for both RR and OR under fully Bayesian method. For both RR and OR we consider $Beta(1, 1)$ in each arm as the non-informative prior. The three values of θ correspond to $f = -0.2, -0.3$ and -0.4 respectively, which correspond to the three choices of the NI margin. This implies that the effect of the experimental drug, must be at least 80%, 70% and 60% respectively of the effect of the active control in order to be non-inferior. From Figure 1, we observe that as θ decreases, the power curve becomes steeper which means for smaller θ the proposed test is more powerful. This is so because for smaller θ it is easier to declare non-inferiority of the experimental drug over the reference.

We refer to the Sections 3.4.1, 3.4.2, and 3.4.3 for the sample size determination under fully Bayesian and two approximation-based Bayesian approaches respectively. We compare the sample size under Bayesian approaches with that obtained under conditional Frequentist approach (Chowdhury et al., 2018b) as briefly described in Section 2. As discussed in Section 2, sample sizes in the placebo, reference, and the experimental arms are denoted by n , $r_1 n$ and $r_2 n$ respectively, with $r_1, r_2 \geq 1$. To compute (n_E, n_R, n_P) , we consider three possible allocations for (E, R, P) : $(1 : 1 : 1)$, $(2 : 2 : 1)$ and $(3 : 2 : 1)$ of the total sample size N . The power expression of both the Frequentist's and the Bayesian normal approximations do not give an explicit solution for n and hence an iterative process is needed. For Frequentist's approach we keep $\alpha = 0.025$, and for Bayesian exact as well as for Bayesian approximation approaches, the sample sizes satisfying $power = 1 - \beta$ also yield estimated type-I error of at most $\alpha = 0.025$. The sample sizes are presented for $\theta = 0.8$ and 0.7 , and for a range of π_E keeping $\pi_R = 0.7$ and $\pi_P = 0.1$, for RR in Table 1 and OR in Table 2. From both the tables we observe that the sample sizes under Bayesian approaches are always smaller than or at most equal to that under Frequentist approach implying the effective gain in sample size using the former.

We present the sample sizes for the placebo arm, and those for the arms R and E can be obtained from the allocation ratios. The total sample size for $(1 : 1 : 1)$ is $3n_P^{(1)}$; that for $(1 : 2 : 2)$ is $5n_P^{(2)}$; while for $(1 : 2 : 3)$ it is $6n_P^{(3)}$, where $n_P^{(1)}$, $n_P^{(2)}$ and $n_P^{(3)}$ are the respective sample size for the placebo arm under the three allocations. From both the tables we observe that the sample size decreases with decrease in θ for a fixed power, which is consistent to the power curve plots that show overpowering the trial with decrease in θ or equivalently increase in the margin. Also this change in sample size with varying θ is robust against the sample size allocations. Although appealing at first glance, one may not want to use a balanced study design from two aspects: (i) due to ethical reasons in case an effective treatment exists, the number of patients receiving the placebo should be kept as small as possible, and (ii) as pointed out by Koch and Tangen (1999), the difference between E and R should be expected to be much smaller than the difference of both of them relative to placebo so that the latter ones are easier to detect. As observed by Pigeot et al. (2003) for continuous outcome, the necessary sample size required for the unbalanced allocations is remarkably smaller

compared to the balanced one. From Table 2 for OR we notice that the necessary sample size is remarkably smaller for the unbalanced allocation (2 : 2 : 1) as compared to a balanced design (1 : 1 : 1) and a minor reduction is again obtained for the unbalanced allocation (3 : 2 : 1) as compared to (2 : 2 : 1). However, for RR the sample sizes do not follow the same pattern as OR with respect to the allocation, as can be seen from Table 1. This might be due to the fact that unlike OR, the logarithm transformation of RR yields a skewed distribution and do not conform well to the normal approximation.

5 Application

We illustrate our proposed Bayesian methods for both RR and OR with a published dataset, described in Higuchi et al. (2009), from a three-arm comparative study on major depressive disorder. We also compare the analysis results with that of the conditional Frequentist method from Chowdhury et al. (2018b). Hida and Tango (2011) as well as Ghosh et al. (2015) also considered this specific dataset in their paper. Hida and Tango (2011) proposed a Frequentist's version of the problem for continuous outcomes while Ghosh et al. (2015) considered the Bayesian version. The objective of the depression trial was to compare the efficacy and safety of duloxetine (E) with those of paroxetine (R) and placebo (P). This study was a double-blinded, randomized, parallel-group active-controlled study of a six-week treatment with the following number of patients in each arm: duloxetine ($n_E = 147$), paroxetine ($n_R = 148$) and placebo ($n_P = 145$). The primary endpoint was continuous type which is the change in HAMD-17 total score from baseline at the end of sixth week. Hida and Tango (2011) considered two binary outcomes for their Frequentist approach namely, Response and Remission. Response is the primary outcome defined as the reduction of more than 50% total. Remission is the secondary outcome which is defined as maintaining HAMD-17 score of ≤ 17 at the end of 6th week. We present the basic data in Table 3, in terms of Response and Remission (see Hida and Tango (2011)). We analyze both the Response and Remission outcomes separately using the conditional Frequentist and fully Bayesian approaches. To make a meaningful interpretation of the effect of the experimental drug, a clinically acceptable margin reflecting the largest loss of effect is chosen to determine non-inferiority of the experimental drug over the control. Here, we vary θ in the range [0.5, 0.8] to explore different possibilities. We use $p^* = 0.975$ to determine NI of duloxetine over paroxetine. For the Frequentist approach we calculate the p -value of the test

for both RR and OR as $p\text{-value} = P_{H_0}(W > W_{\text{obs}}) = 1 - \Phi\left(\frac{W_{\text{obs}} - \mu_w^{\text{null}}}{\sigma_w^{\text{null}}}\right)$, where W_{obs} is the

conditional Frequentist's test statistic and μ_w^{null} and $(\sigma_w^{\text{null}})^2$ are the mean and variance of W under H_0 (Chowdhury et al., 2018b). Now for the analysis under fully Bayesian approach, we assume $\text{Beta}(1, 1)$ as the non-informative prior for $\pi_i, i \in \{E, R, P\}$ for both RR and OR, and generate the posterior samplers from Beta distributions as given in Step 3 of simulation. We calculate the Bayesian posterior probability as given in Step 4 of simulation. The Frequentist p -values and the Bayesian posterior probabilities $P(H_1|Data)$ are reported in Table 4 and Table 5 for RR and OR respectively for both the Response and the Remission data.

The Frequentist's p -values are compared with $\alpha = 0.025$ while the Bayesian posterior probabilities are compared with p^* to deduce the respective decision. The decision (claim NI or not) is reported for both the Frequentist and Bayesian approaches. Here "1" stands for rejection of H_0 implying NI is claimed while "0" represents the acceptance of H_0 . From both Table 4 and Table 5 we observe that the posterior probabilities increase as θ decreases implying greater chance of declaring NI for smaller values of θ , which is consistent with the simulation results. Similar is the conclusion for the Frequentist's p -values which increase as θ increases, since p -value $< \alpha$ implies rejection of NI. From both the tables we observe that, under the non-informative prior, the posterior probabilities for both the Response and Remission data are less than the pre-specified cut-off $p^* = 0.975$ and hence non-inferiority of E relative to R cannot be claimed. However, when we choose an appropriate informative Beta prior, NI is established for smaller values of θ . For both RR and OR we choose the informative Beta priors for the three arms as E : $Beta(40, 34.22)$, R : $Beta(40, 35.584)$ and P : $Beta(40, 64.63)$ under which NI is established for $\theta = 0.55$ for the Response data. For the Remission data we assume the following informative priors: E : $Beta(40, 76.7)$, R : $Beta(40, 80.19)$ and P : $Beta(40, 139.27)$ and that establishes NI for $\theta = 0.5$. Note that the informative priors are so chosen that the mode of the specified Beta distributions equal the parameter estimates from the data: $\hat{\pi}_E = 0.54$, $\hat{\pi}_R = 0.53$ and $\hat{\pi}_P = 0.38$ for the Response data and $\hat{\pi}_E = 0.34$, $\hat{\pi}_R = 0.33$ and $\hat{\pi}_P = 0.22$ for the Remission data. Also, note that other than equating the mode of Beta distribution with the parameter estimates, the parameter values are arbitrarily chosen for illustration purpose, taking into account the fact that the variances of these informative Beta distributions are smaller than that of the non-informative $Beta(1, 1)$ distribution.

6 Discussion

In this paper, we have presented new Bayesian test procedures for the "gold standard" three-arm NI trial, which includes a placebo arm for binary endpoints, considering RR and OR as the functionals of interest. We believe that this is an important methodological contribution as NI can be claimed based on these functionals according to the recent FDA guideline. In the fully Bayesian approach, we explored standard conjugate prior and carried out the simulation considering non-informative priors. We also proposed two approximation-based Bayesian approaches which avoid substantial part of the computation of the fully Bayesian approach. We tabulated the sample size in Tables 1 and 2 under three different types of allocation for both RR and OR. This should provide a good starting point for accessing sample size and allocation requirement when designing such trials. We have seen that even with non-informative prior, fully Bayesian approach as well as Bayesian approximation-based approaches yield smaller sample size as compared to Frequentist's approach for a desired power of 80%. The sample size gain is substantial for informative prior choices. Also analysis of our real clinical trial data suggests that the Bayesian methods perform favorably in all situations. Note that, the historical information plays substantial role in the design and analysis of NI trial. Hence NI trial has to be reflected in several substantive aspects including the choice of δ , the question of whether a placebo can be included as an additional arm of the study, assay sensitivity, to give a few examples among others.

We also noted that under the fraction margin approach the fraction “ f ” is pre-specified, while the NI margin δ is unknown. Hence, the value of δ can vary greatly depending on the estimated effect size of the reference treatment, i.e. as a function of $(g(\hat{\pi}_R), g(\hat{\pi}_P))$. On the other hand, in the fixed margin approach (see, Hida and Tango (2011) and Ghosh et al. (2015)), with the three-arms, the joint testing of NI and AS may be performed which needs additional care since it may produce conservative decision and may yield restrictive type-I error (Chuang-Stein et al., 2007; Dmitrienko et al., 2009). Albeit, development of such procedure for RR and OR under alternative definition of type-I error (e.g., average testing error of Chuang-Stein et al. (2007)) and under Bayesian calibration (Ghosh et al., 2015) is an interesting future problem.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The research of last author is partly supported by PCORI contract number ME-1409-21410 and NIH grant number P30-ES020957. Authors would also like to thank two anonymous referees, AE and the editorial team, whose comments provided additional insights and have greatly improved the scope and presentation of the paper.

References

- (2005). Guideline on the choice of the noninferiority margin (Doc. Ref. EMEA/CPMP/EWP/215) EMA.
- (2016). Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry FDA.
- Althunian TA, de Boer A, Klungel OH, Insani WN, and Groenwold RH (2017). Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. *Trials*, 18(1):107. [PubMed: 28270184]
- Arnold BC and Beaver RJ (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58(3):471–488.
- Brown D, Volkens P, and Day S (2006). An introductory note to chmp guidelines: choice of the non-inferiority margin and data monitoring committees. *Statistics in medicine*, 25(10):1623–1627. [PubMed: 16639776]
- Chowdhury S, Tiwari R, and Ghosh S (2018a). Approaches for testing non-inferiority in two-arm trial for risk ratio and odds ratio. *Journal of Biopharmaceutical Statistics*, Accepted for Publication
- Chowdhury S, Tiwari RC, and Ghosh S (2017). Non-inferiority testing for three-arm trials with binary outcome: Novel frequentist and bayesian proposals. Under review
- Chowdhury S, Tiwari RC, and Ghosh S (2018b). Non-inferiority testing for risk ratio, odds ratio and number needed to treat in three-arm trial. *Computational Statistics & Data Analysis* Available online first 15 September 2018.
- Chuang-Stein C, Stryzszak P, Dmitrienko A, and Offen W (2007). Challenge of multiple co-primary endpoints: a new approach. *Statistics in medicine*, 26(6):1181–1192. [PubMed: 16927251]
- D’Agostino RB, Massaro JM, and Sullivan LM (2003). Noninferiority trials: Design concepts and issues-the encounters of academic consultants in statistics. *Statistics in Medicine*, 22(2):169–186. [PubMed: 12520555]
- Dmitrienko A, Tamhane AC, and Bretz F (2009). Multiple testing problems in pharmaceutical statistics CRC Press.
- Gamalo MA, Tiwari RC, and LaVange LM (2014). Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. *Pharmaceutical Statistics*, 13(1):25–40. [PubMed: 23913880]

- Gamalo MA, Wu R, and Tiwari RC (2011). Bayesian approach to noninferiority trials for proportions. *Journal of biopharmaceutical statistics*, 21(5):902–919. [PubMed: 21830922]
- Gamalo MA, Wu R, and Tiwari RC (2012). Bayesian approach to non-inferiority trials for normal means. *Statistical Methods in Medical Research*, 25(1):221–240. [PubMed: 22619277]
- Ghosh P, Nathoo F, Gonen M, and Tiwari RC (2011). Assessing noninferiority in a three-arm trial using the bayesian approach. *Statistics in Medicine*, 30(15):1795–1808. [PubMed: 21520456]
- Ghosh S, Ghosh S, and Tiwari RC (2015). Bayesian approach for assessing non-inferiority in a three-arm trial with pre-specified margin. *Statistics in medicine*, 35(5):695–708. [PubMed: 26434554]
- Ghosh S, Tiwari RC, and Ghosh S (2018). Bayesian approach for assessing non-inferiority in a three-arm trial with binary endpoint. *Pharmaceutical Statistics*, 17(4):342–357. [PubMed: 29473291]
- Hashemi L, Nandram B, and Goldberg R (1997). Bayesian analysis for a single 2 by 2 table. *Statistics in Medicine*, 16(12):1311–1328. [PubMed: 9232755]
- Hida E and Tango T (2011). On the three-arm noninferiority trial including a placebo with a prespecified margin. *Statistics in Medicine*, 30(3):224–231. [PubMed: 21213340]
- Higuchi T, Murasaki M, and Kamijima K (2009). 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.
- Hilton JF (2010). Noninferiority trial designs for odds ratios and risk differences. *Statistics in Medicine*, 29(9):982–993. [PubMed: 20191595]
- Hung HMJ and Wang SJ (2004). Multiple testing of noninferiority hypotheses in active controlled trials. *Journal of Biopharmaceutical Statistics*, 14(2):327–335. [PubMed: 15206530]
- ICHE10 (2009). 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.
- ICHE9 (2009). ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.
- Kieser M and Friede T (2007). Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in medicine*, 26(2):253–273. [PubMed: 16685703]
- Kieser M and Stucke K (2016). Assessing additional benefit in noninferiority trials. *Biometrical Journal*, 58(1):154–169. [PubMed: 24979706]
- Koch GG and Tangen CM (1999). Non parametric analysis of covariance and its role in non-inferiority clinical trials. *Drug information journal*, 33(4):1145–1159.
- Pigeot I, Schafer J, Rohmel J, and Hauschke D (2003). Assessing noninferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22(6):883–899. [PubMed: 12627407]
- Rousson V and Seifert B (2008). A mixed approach for proving non-inferiority in clinical trials with binary endpoints. *Biometrical Journal*, 50(2):190–204. [PubMed: 18311852]
- Schumi J and Wittes JT (2011). Through the looking glass: understanding non-inferiority. *Trials*, 12(2):106–118. [PubMed: 21539749]
- Wange G, Roes KCB, Boer AD, Hoes AW, and Knol MJ (2013). The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants. *Canadian Medical Association Journal*, 185(3):222–227. [PubMed: 22908144]

Figure 1:
Power curves for different θ for RR test in (a) and OR test in (b).

Table 1:

Sample size for RR under Frequentist, fully Bayesian and Bayesian Approximation approaches to achieve a power of 80% for $\theta = 0.8$ and 0.7 , $\alpha = 0.025$ and $\pi_E \in [0.65, 0.9]$ under three different allocations

Allocation			θ	π_E	Frequentist		Fully Bayesian		Prior Approx		Posterior Approx.	
P	R	E			n_P	N	n_P	N	n_P	N	n_P	N
1	1	1	0.8	0.9	27	81	23	69	22	66	22	66
				0.85	33	99	28	84	28	84	28	84
				0.8	42	126	37	111	36	108	36	108
				0.75	56	168	51	153	49	147	51	153
				0.7	79	237	71	213	73	219	73	219
				0.65	124	372	118	354	118	354	119	357
	1	1	0.7	0.9	24	72	17	51	18	54	17	51
				0.85	28	84	20	60	21	63	20	60
				0.8	33	99	25	75	26	78	26	78
				0.75	40	120	33	99	34	102	34	102
				0.7	51	153	45	135	45	135	43	129
				0.65	68	204	60	180	61	183	62	186
1	2	2	0.8	0.9	17	85	13	65	14	70	13	65
				0.85	21	105	17	85	19	95	18	90
				0.8	27	135	26	130	24	120	25	125
				0.75	35	175	33	165	32	160	32	160
				0.7	49	245	46	230	46	230	48	240
				0.65	76	380	76	380	74	370	76	380
	2	2	0.7	0.9	17	85	11	55	15	75	12	60
				0.85	19	95	14	70	17	85	15	75
				0.8	23	115	19	95	22	110	19	95
				0.75	28	140	24	120	26	130	23	115
				0.7	35	175	33	165	35	175	32	160
				0.65	47	235	43	215	45	225	46	230
1	2	3	0.8	0.9	15	90	10	60	13	78	11	66
				0.85	18	108	13	78	16	96	14	84
				0.8	23	138	19	114	22	132	20	120
				0.75	30	180	24	144	27	162	26	156
				0.7	42	252	37	222	39	234	40	240
				0.65	64	384	62	372	64	384	64	384
	3	3	0.7	0.9	14	84	8	48	12	72	9	54
				0.85	17	102	9	54	15	90	12	72
				0.8	20	120	14	84	19	114	15	90

Allocation			θ	π_E	Frequentist		Fully Bayesian		Prior Approx		Posterior Approx.	
P	R	E			n_P	N	n_P	N	n_P	N	n_P	N
				0.75	24	144	17	102	22	132	18	108
				0.7	31	186	25	150	28	168	25	150
				0.65	41	246	33	198	37	222	35	210

Table 2:

Sample size for OR under Frequentist, fully Bayesian and Bayesian Approximation approaches to achieve a power of 80% for $\theta = 0.8$ and 0.7 , $\alpha = 0.025$ and $\pi_E \in [0.65, 0.9]$ under three different allocations

Allocation			θ	π_E	Freq		Fully Bayesian		Prior Approx.		Posterior Approx.	
P	R	E			n_P	N	n_P	N	n_P	N	n_P	N
1	1	1	0.8	0.9	20	60	20	60	20	60	20	60
				0.85	31	93	25	75	27	81	27	81
				0.8	49	147	45	135	46	138	47	141
				0.75	85	255	82	246	82	246	83	249
				0.7	165	495	160	480	157	471	158	474
				0.65	415	1245	404	1212	406	1218	406	1218
			0.7	0.9	15	45	14	42	14	42	14	42
				0.85	21	63	15	45	17	51	18	54
				0.8	30	90	26	78	28	84	28	84
				0.75	45	135	39	117	40	120	42	126
				0.7	72	216	65	195	62	186	63	189
				0.65	125	375	116	348	114	342	116	348
1	2	2	0.8	0.9	11	55	11	55	11	55	11	55
				0.85	16	80	15	75	15	75	16	80
				0.8	26	130	24	120	26	130	26	130
				0.75	45	225	40	200	43	215	43	215
				0.7	87	435	82	410	83	415	83	415
				0.65	220	1100	212	1060	211	1055	212	1060
			0.7	0.9	8	40	8	40	8	40	8	40
				0.85	12	60	9	45	11	55	11	55
				0.8	17	85	16	80	16	80	17	85
				0.75	25	125	21	105	22	110	23	115
				0.7	41	205	36	180	35	175	36	180
				0.65	71	355	62	310	63	315	64	320
1	2	3	0.8	0.9	9	54	8	48	8	48	8	48
				0.85	13	78	12	72	12	72	12	72
				0.8	22	126	19	114	21	126	20	120
				0.75	37	222	33	198	35	210	36	216
				0.7	72	432	70	420	68	408	70	420
				0.65	182	1092	177	1062	175	1050	177	1062
			0.7	0.9	7	42	6	36	6	36	6	36
				0.85	10	60	9	54	9	54	9	54
				0.8	14	84	11	66	13	78	13	78

Allocation			θ	π_E	Freq		Fully Bayesian		Prior Approx.		Posterior Approx.	
P	R	E			n_P	N	n_P	N	n_P	N	n_P	N
				0.75	22	126	17	102	18	108	19	114
				0.7	34	204	30	180	29	174	30	180
				0.65	60	360	54	324	52	312	53	318

Table 3:

Remission and Response as Outcome in the Depression Trial of Higuchi et al. (2009)

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

Table 4:Frequentist p -values and Bayesian posterior probabilities, and rejection decision for Risk Ratio

θ	Response						Remission					
	Freq	Dec	Bayes Non-inform	Dec	Bayes Inform	Dec	Freq	Dec	Bayes Non-inform	Dec	Bayes Inform	Dec
0.5	0.0514	0	0.971	0	0.983	1	0.1055	0	0.917	0	0.979	1
0.55	0.0650	0	0.953	0	0.977	1	0.1249	0	0.898	0	0.971	0
0.6	0.0826	0	0.938	0	0.968	0	0.1483	0	0.874	0	0.962	0
0.65	0.1052	0	0.922	0	0.946	0	0.1760	0	0.858	0	0.942	0
0.7	0.1335	0	0.893	0	0.932	0	0.2082	0	0.826	0	0.91	0
0.75	0.1681	0	0.853	0	0.908	0	0.2450	0	0.786	0	0.879	0
0.8	0.2094	0	0.814	0	0.86	0	0.2861	0	0.738	0	0.821	0

Table 5:Frequentist p -values and Bayesian posterior probabilities, and rejection decision for Odds Ratio

θ	Response						Remission					
	Freq	Dec	Bayes Non-inform	Dec	Bayes Inform	Dec	Freq	Dec	Bayes Non-inform	Dec	Bayes Inform	Dec
0.5	0.0462	0	0.97	0	0.985	1	0.1025	0	0.909	0	0.979	1
0.55	0.0622	0	0.95	0	0.979	1	0.1239	0	0.884	0	0.974	0
0.6	0.0829	0	0.926	0	0.96	0	0.1496	0	0.857	0	0.96	0
0.65	0.1091	0	0.901	0	0.942	0	0.1795	0	0.826	0	0.941	0
0.7	0.1410	0	0.868	0	0.922	0	0.2138	0	0.791	0	0.913	0
0.75	0.1788	0	0.831	0	0.885	0	0.2520	0	0.758	0	0.88	0
0.8	0.2218	0	0.777	0	0.757	0	0.2935	0	0.716	0	0.847	0