

Assessment of Noninferiority (and Equivalence) for Simple Crossover Trials Using Bayesian Approach

Mingan Yang¹

Received: 5 December 2016 / Accepted: 28 November 2017 / Published online: 17 January 2018
© International Chinese Statistical Association 2018

Abstract To assess the noninferiority or equivalence of a general drug to a standard one, researchers generally use the odds ratio of patient response rates to evaluate the relative treatment efficacy. In this paper, we use a logistic random effects model to test noninferiority and equivalence based on the odds ratio of patient response rates for crossover trials with binary data. We use Bayesian sampling algorithm, data augmentation, and scaled mixture of normal representation to implement the approach and improve efficiency. The performance of the proposed approach is assessed via simulation and real data examples.

Keywords Binary data · Crossover trial · Equivalence · Noninferiority · Odds ratio

1 Introduction

In medical studies, researchers are generally interested in testing whether a new drug is not worse than a standard drug. Generally, the efficacy is not the only factor considered. The new drug is assumed beneficial if its efficacy is not inferior to the standard one by some prespecified standards if the new drug is proved to have advantages such as less toxicity, low cost, or easier administration. However, some drugs do not produce any appreciable absorption into the systemic circulation, and thus, the pharmacokinetics approaches are not appropriate in assessing noninferiority or bioequivalence [1–6]. Instead, the clinical endpoints are used directly to assess the equivalence.

✉ Mingan Yang
myang@mail.sdsu.edu

¹ Division of Biostatistics and Epidemiology, Graduate School of Public Health, San Diego State University, San Diego, CA, USA

Although researchers often use the risk difference and risk ratio as measures of risk for comparing different drugs or treatments in prospective clinical trials, they generally can not be used to assess noninferiority or equivalence of risk or disease outcomes in retrospective studies. On the other hand, odds ratio is widely used as a measure of risk association in both prospective and retrospective studies.

In clinical trials, a crossover design is widely used to assess the drug effects. The most commonly used crossover design is a design of two treatments A and B, in which some subjects receive treatment A first and B second, while the others receive treatment B first and A second. One of the main benefits of cross trials is that the inherent variability is reduced, and hence the power is increased [7]. Medical researchers also use crossover trials to assess bioequivalence of drugs [2, 8]. It often involves comparisons of the response rates of binary outcomes to test for noninferiority or equivalence between two drugs. For example, in comparing efficacy of two drugs, researchers study the therapeutic response of relief of primary dysmenorrhea (yes = 1, no = 0) in the crossover trial. In this article, we focus on studies where the patient response outcome is dichotomous. That is, we want to assess whether the positive response rate for the generic drug is not inferior to that of the standard drug by a prespecified margin. The test is generally denoted as noninferiority trial [9].

In crossover design, it is generally assumed that the carryover effect, the effect of a drug that persists after the end of the dosing period, does not exist with a sufficiently long washout period. However, if the assumption is not valid, it is not appropriate to use the crossover design since it causes the biased estimate of the treatment effect. [10, 11]. In practice, we may always be able to use the crossover trial with a sufficient washout period in between administering drugs under comparison. There are many papers on crossover trials in current literatures. Gart [12] used a logistic regression model to test the equality of two treatment effects. Zimmermann and Rahlfs [13] used the linear additive risk model to test equality in patient response rates for a simple crossover trial. Schouten and Kester [14] used the similar approach to assess treatment effects for crossover trials. Ezzet and Whitehead [15] used a logistic random effect to assess equality of the treatment effects. Lui and Chang [16] proposed a semiparametric approach to assess the equality test. Senn [17, 18] provided a general overview of the current progress in crossover design. On the other hand, there are very limited Bayesian approaches in the literature. Osman and Ghosh [19] proposed a Bayesian semiparametric approach to test noninferiority. Ghosh et al. [20] proposed a Bayesian semiparametric approach to test noninferiority in three-arm trials. Although these approaches generally work well, there are some obvious disadvantages. For example, both approaches of Ghosh et al. [20], and Osman and Ghosh [19] are not easily implemented and do not characterize heterogeneity among the subjects. The approach of Lui and Chang [16] is subject to asymptotic theorem constraint and does not work well with small data sets. In particular, the approach of Lui and Chang [16] does not work when some cells have zero values. To the author's best knowledge, there is no Bayesian approach proposed to test noninferiority or equivalence based on the odds ratio of patient response rates for a crossover design.

In this paper, we propose a Bayesian approach based on the odds ratio of patient response rate to test noninferiority and equivalence for simple crossover designs. Compared with the frequentist approaches, a fully Bayesian approach can incorporate

useful prior information to account for varied uncertainties. Statistical inference can be easily derived efficiently and accurately using Markov chain Monte Carlo algorithms from the posterior distributions. Thus, the results of the hypotheses tests of noninferiority (or equivalence) do not suffer from the asymptotic constraints. In addition, the Bayesian approach generally has advantage for hypothesis testing since it is based on the posterior probability distribution instead of the p values that are often misinterpreted.

The paper is organized as follows: Sect. 2 describes the model. Section 3 provides a simulation study to demonstrate the performance of the proposed approach. Section 4 illustrates the approach using a real data example, followed by conclusion and discussion in Sect. 5.

2 The Model

2.1 General Description

We consider a simple crossover trial for the noninferiority (or equivalence) test of a new drug B versus a standard drug A. Assume that n_1 patients are assigned to the first group ($g = 1$), for which the patients take drug A and B at period 1 and 2, respectively; n_2 patients assigned to the second group ($g = 2$) for which patients take drug B and A at period 1 and 2, respectively. We assess the treatment efficacy by conducting noninferiority (or equivalence) test based on the odds ratio of positive response rate of the patients. Let y_{gij} denote the patient binary outcome from subject i ($i = 1, \dots, n_g$) at period j ($j = 1, 2$) from treatment sequence g ($g = 1, 2$); the model is specified as follows:

$$y_{gij} \sim \text{Bernoulli}(H^{-1}(\chi_{gij})), \quad \chi_{gij} = \vartheta_{gj} + \tau_{g(1,2)}I(j = 2) + \mu_{gi}, \quad (1)$$

where $\vartheta_{gj} = u_0 + \eta_g + \psi_j + t_{l(g,j)}$ and u_0 is the overall mean; η_g is the fixed effect of the g th group sequence, $g = 1, 2$; ψ_j is the fixed effect of the j th period; $t_{l(g,j)}$ is the fixed treatment effect, and $l(g, j)$ is the treatment index; $\tau_{g(1,2)}$ is the first-order carryover effect of treatment at period 1 from the immediately preceding period into the effect of treatment at period 2 in the current period; μ_{gi} is the random effect of i th subject from the g th sequence, $i = 1, \dots, n_g$, specified as a normal distribution with $\mu_{gi} \sim N(0, \rho^2)$ and $H(\cdot)$ is the logistic link function with $H(\kappa) = \log(\kappa/(1 - \kappa))$. We set $y_{giz} = 1$ if the patient has the positive response outcome of interest (with improvement), and 0 (without improvement) otherwise. We impose the following constraints for the group effect, treatment effect, and period effect $\sum \eta_g = 0$, $\sum \psi_j = 0$, $\sum_{m=1}^M t_m = 0$ to resolve identifiability.

The carryover effect is the effect of a drug that persists after the end of the dosing period. It poses an challenging issue in analyzing the crossover trials since it causes the biased estimate of the treatment effect. If there is a sufficient washout between dosing periods, then the carryover effect can be ignored. Obviously, it is impossible to get all the parameters in the above formula given limited information. In this article, we will focus mainly on the treatment effects and drop out the group sequence effect and

the carryover effect, assuming sufficient long washout in between two dosing periods. Thus, model specified in Eq. (1) is reduced to

$$y_{gij} \sim \text{Bernoulli}(H^{-1}(\chi_{gij})), \quad \chi_{gij} = u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}. \tag{2}$$

Although the carryover effect is removed in the above model, we should note that it can be easily estimated following general procedures, e.g., [21,22]. Obviously, the logistic model is nonlinear and we can not get conditional conjugacy even with simple normal priors, which ultimately causes inefficiency in computation. We will take several approaches to improve efficiency by converting the nonlinear model to the standard linear models. According to the current literature [23–25], the logistic distribution can be approximated by the Student’s t distribution. The Student’s t distribution can be expressed as a scale mixture of normals [26]. Thus, model specified in Eq. (2) can be expressed equivalently as follows with auxiliary variables:

$$y_{gij} = 1 : y_{gij}^* > 0, \\ y_{gij} = 0 : y_{gij}^* \leq 0,$$

where y_{gij}^* is an underlying value with the logistic distribution with location parameter $u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}$ and density function as follows:

$$f(y_{gij}^*|\cdot) = \frac{\exp\{- (y_{gij}^* - (u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}))\}}{\{1 + \exp[- (y_{gij}^* - (u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}))]\}^2}. \tag{3}$$

Thus, y_{gij}^* is approximated as a noncentral t distribution with location parameter $u_0 + \eta_g + \psi_j + t_{l(g,j)} + \mu_{gi}$, degree of freedom v , and scale parameter σ^2 . We can further express it as a scale mixture of normals and get the following model:

$$y_{gij}^* = u_0 + \eta_g + \psi_j + t_{l(g,j)} + \mu_{gi} + \epsilon_{gij}, \quad \epsilon_{gij} \sim N(0, \sigma^2/\phi_{gij}), \tag{4}$$

where ϕ_{gij} has a Gamma prior $G(v/2, v/2)$. We take $v = 7.3$ and $\sigma^2 = \pi^2(v - 2)/3v$ as suggested by O’Brien and Dunson [25] to make the approximation almost exact.

2.2 Prior and Posterior

A prior specification for unknown parameters is essential and important for Bayesian approaches. It is recommended not to use excessively diffuse or flat priors since they might cause improper posterior due to the intractable nature of the density. On the other hand, it is preferred to choose conjugate priors to facilitate computation and improve efficiency. We follow these rules and choose priors with caution in our Bayesian approach.

We specify a normal distribution $N(\mu_1, \sigma_1^2)$ for the overall mean $u_0 \sim N(\mu_1, \sigma_1^2)$. Similar priors are selected for the period effect and the treatment effect: $\psi_j \sim N(\mu_3, \sigma_3^2)$, and $t_m \sim N(\mu_4, \sigma_4^2)$. The random effect μ_{gi} is specified as $\mu_{gi} \sim$

$N(0, \rho^2)$; the hyperparameter ρ^2 is placed an Inverse Gamma distribution $\rho^2 \sim IG(a_0, b_0)$, and ϕ_{gij} is placed a prior of Gamma distribution $G(v/2, v/2)$. Based on the model and prior specifications, we can easily derive the joint posterior distribution for $\tilde{\theta} = (u_0, \psi, \mathbf{t}, \phi)$ as follows:

$$p(\tilde{\theta}|\mathbf{y}) \propto p(\cdot) \left[\prod_g \prod_i N(\mu_{gi}; 0, \rho^2) \prod_j N(y_{gij}; u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}, \sigma^2/\phi_{gij}) w_{gij} \right], \tag{5}$$

where $w_{gij} = \{1(y_{gij}^* > 0)y_{gij} + 1(y_{gij}^* < 0)(1 - y_{gij})\}p(\phi_{gij})$, and $p(\cdot) = p(\rho^2)p(u_0)p(\mathbf{t})p(\psi)$. Obviously, we get a very complicated posterior formula that we can not sample directly. By introducing the latent variable y_{gij}^* , we have applied a data augmentation algorithm and can easily sample the parameters and hyperparameters of interest using Gibbs sampler. The main idea of data augmentation algorithm is to augment the observed data Y with another variable V , which is generally referred to as latent data. Given Y and V , one can easily sample the parameters θ from the posterior distribution $P(\theta|Y, V)$. From the above model, the auxiliary variable can be easily updated using the Gibbs sampler from a posterior normal distribution truncated below or above 0 according to the value of y_{gij} . The conditional posterior of the auxiliary variable is

$$p(y_{gij}^*|\theta, y_{gij}) = \frac{N(y_{gij}; q_{gij}, \sigma^2/\phi_{gij})\{1(y_{gij}^* > 0)y_{gij} + 1(y_{gij}^* < 0)(1 - y_{gij})\}}{\Phi(0; q_{gij}, \sigma^2/\phi_{gij})^{1-y_{gij}}\{1 - \Phi(0; q_{gij}, \sigma^2/\phi_{gij})\}^{y_{gij}}}, \tag{6}$$

where $q_{gij} = u_0 + \psi_j + t_{l(g,j)} + \mu_{gi}$. The full conditional posterior distribution is specified in (5). The detailed sampling steps are listed in the Appendix. We run the Gibbs sampler by iteratively sampling all the parameters, and hyperparameters of interest.

Given the full conditional posterior distributions in formula (5), one can easily use the MCMC algorithms to derive the estimates of the parameters of interest. After discarding the initial burn-in period, we can get the posterior summaries of parameters of interest from the Gibbs sampler output. For example, if we want to evaluate the relative treatment effect difference between drug B and drug A. The posterior estimated parameter t_{BA} is given

$$t_{BA} = \sum_{k=\zeta+1}^K (t_B^{(k)} - t_A^{(k)}) / (K - \zeta), \tag{7}$$

where the relative treatment effect difference is defined as $t_{BA} = t_B - t_A$, $t_B^{(k)}$ is the k th draw of t_B from the posterior sampling. We denote $v_{BA} = \exp(t_{BA})$. Obviously v_{BA} represents the odds ratio of relative difference between treatment A and B. The relative

treatment effect difference derived in this section is used to assess noninferiority and equivalence of drug B (treatment B) to drug A (treatment A) in relation to treatment efficacy, to be discussed in the next section.

2.3 Test Noninferiority and Equivalence

Generally, the noninferiority hypothesis of between two treatments is formulated

$$H_0 : t_B - t_A \leq \omega \quad \text{vs} \quad H_1 : t_B - t_A > \omega, \quad (8)$$

where $\omega < 0$ denotes the predetermined real value quantity, and is called the amount of noninferiority margin. Alternatively, we can also use the following formula based on the odds ratio of the response rate:

$$H_0 : \nu_{BA} \leq \delta \quad \text{vs} \quad H_1 : \nu_{BA} > \delta, \quad (9)$$

where $0 < \delta < 1$. When the null hypothesis is rejected, the conclusion of noninferiority is reached for the experimental treatment (B) to the reference treatment (A).

The formula of noninferiority testing can be modified easily for hypothesis test of equivalence between two treatments. The hypothesis test of equivalence between two treatments B and A is formulated as

$$H_0 : \nu_{BA} \leq \delta_1 \quad \text{or} \quad \nu_{BA} \geq \delta_2 \quad \text{vs} \quad H_a : \delta_1 < \nu_{BA} < \delta_2 \quad (10)$$

where δ_1 and δ_2 are the predetermined maximum clinical acceptable margins. When the null hypothesis is rejected, one can arrive at the conclusion of equivalence between two treatments.

Given the full conditional posterior distributions, we can easily get the posterior probabilities of the hypothesis using the MCMC algorithm from the posterior draws by following (7) in a similar manner.

3 Simulation

To evaluate the performance of our approach, we conduct a simulation study. The specification of Bayesian model is completed by specifying the prior. Basu and Santra [21] chose fairly flat but proper, conditionally conjugate priors to analyze crossover trial studies. Following this line, we chose similar priors for our analysis. We assume that there are no carry-over effects with an adequate washout period for the simulation. The overall mean u_0 is set equal to 0.10. We generate the random effects μ_{gi} independently and identically from a normal distribution with mean 0 and standard deviation $s = 0.2, 0.5, \text{ and } 0.8$, respectively; we set the relative treatment effect difference t_{21} as 0.1, 0.3, and 0.5, respectively. The number of patients for each group is varied as $n_1 = n_2 = 15, 20, 50, \text{ and } 150$. For the priors, we specify $\mu_1 = \mu_2 = \mu_3 = 0.5$, $\sigma_1 = \sigma_2 = \sigma_3 = 10.0$, and $a_0 = b_0 = 0.1$.

We generate 200 simulated data sets according to Eq. (2). The percentage of data sets with zero count cells varies with different scenarios. For example, for small sample size $n_1 = n_2 = 15$, there are about 5% ($s = 0.2$) to 12% ($s = 0.8$) data sets with zero counts, whereas for $n_1 = n_2 = 20$, there are about 2% to 4% data sets with zero counts. We take the Gibbs sampling algorithm as described in the previous section and the Appendix. The Gelman–Rubin approach [27] is utilized to assess the mixing and convergence. After an initial 2000 iterations, we take the next 5000 iterations to estimate the parameters of interest. We also run the simulations with varying means and variances of the priors for the parameters to evaluate the effects. For data sets of small sample size, the length of the coverage is a little wider with flatter prior, but negligible with bigger sample size. We generally do not see any significant deviation in the parameter estimation.

We compare our proposed approach with that of Lui and Chang [16]. For easy notation, we denote our approach as YB, Lui and Chang [16] as LC. The program is compiled in Intel Visual Fortran Professional 16.0 and executed in a PC (Win7 Professional, Intel Xeon processor E5-1620 at 3.5 GHz, 6 G RAM). We ran the Gibbs sampler for above simulation. Although we make a few approximation, the program runs very fast, e.g., the CPU time consumed is 642s for $n = 50$ with 200 simulated data sets. Table 1 provides the bias, mean square error (MSE), and the 95% coverage of the relative treatment effect difference of varying scenarios. From the table, we can see that our approach (YB) has fairly smaller bias and MSE than that of LC, though a little smaller in coverage for small sample size. As the sample size increases, the coverage and MSE of YB are still better than those of LC, and the bias of both approaches is very close to each other. This might be due to the presence of zero count cells for small sample size, which is difficult for the approach LC to estimate. As the sample size increases, the results of LC get improved since there is less likely presence of zero count cells. Overall, we can see the performance of our approach is very good. Although there are some data sets that have cells with zero count value, a challenging issue for the approach of Lui and Chang [16], our approach does not suffer from it and can still provide very reliable and consistent results.

We also study the relationship between the posterior probability of H_1 and the relative treatment effect difference in the above simulations. Figure 1 provides the results. Within each panel of Fig. 1, the posterior probabilities of the alternative hypothesis are in ascending order as the standard deviation of the residual errors increases at the beginning, and in descending order at later time. The posterior probability of H_1 increases as the relative treatment effect difference increases. When we check across the panels, we see that the curve becomes steeper as the sample size increases. These results are expected and are in accordance with intuition.

4 Applications

In this section, we use two real data examples to illustrate our approach. For the first real data example in Table 2, we consider the data set given by Senn [17]. In this study, 24 children who suffered from exercise-induced asthma were randomly assigned to a crossover trial of two treatment sequences: drug B, 12 μg formoterol solution aerosol;

Table 1 The estimated bias, MSE, and 95% coverage of the relative treatment effect difference

<i>n</i>	<i>t</i> ₂₁	σ	YB			LC		
			Bias	MSE	Coverage	Bias	MSE	Coverage
15	0.5	0.8	3.76e−2	0.31	0.97	−3.27e−2	0.44	0.98
		0.5	4.12e−2	0.25	0.960	−1.12e−2	0.34	0.99
		0.2	6.24e−2	0.32	0.94	−8.91e−2	0.41	0.99
	0.3	0.8	3.49e−2	0.24	0.96	−8.12e−2	0.34	0.97
		0.5	3.85e−2	0.32	0.95	−6.45e−2	0.43	0.97
		0.2	5.78e−2	0.34	0.94	−6.53e−2	0.41	0.97
	0.1	0.8	1.74e−2	0.29	0.95	−3.57e−2	0.43	0.98
		0.5	8.05e−4	0.33	0.94	−3.64e−3	0.41	0.96
		0.2	8.45e−3	0.359	0.93	1.26 e−2	0.41	0.97
20	0.5	0.8	−2.43e−2	0.22	0.95	−6.96e−2	0.32	0.97
		0.5	−2.79e−2	0.238	0.94	−1.32e−2	3.13e−1	0.96
		0.2	−1.19e−2	0.231	0.94	−3.25e−2	2.89e−1	0.95
	0.3	0.8	−1.06e−2	1.96e−1	0.96	−5.84e−2	3.12e−1	0.97
		0.5	1.09e−2	2.13e−1	0.96	−5.61e−2	2.97e−1	0.96
		0.2	1.56e−2	2.37e−1	0.94	−3.10e−2	2.91e−1	0.94
	0.1	0.8	2.81e−3	1.98e−1	0.96	−3.19e−2	2.98e−1	0.97
		0.5	1.45e−2	2.25e−1	0.95	−3.01e−2	2.94e−1	0.96
		0.2	1.15e−2	2.19e−1	0.94	−7.45e−3	2.78e−1	0.96
50	0.5	0.8	−4.35e−2	8.27e−2	0.96	−3.09e−2	1.21e−1	0.95
		0.5	−9.85e−3	8.41e−2	0.95	−2.78e−2	1.03e−1	0.96
		0.2	1.30e−2	7.87e−2	0.95	−2.67e−2	7.96e−2	0.96
	0.3	0.8	−3.80e−2	7.61e−2	0.97	−1.52e−2	1.01e−1	0.97
		0.5	−1.11e−2	8.54e−2	0.95	−1.21e−2	1.02e−1	0.95
		0.2	−6.34e−3	8.79e−2	0.95	−1.64e−2	9.67e−2	0.945
	0.1	0.8	−1.04e−2	7.74e−2	0.96	−6.16e−3	1.05e−1	0.96
		0.5	5.07e−3	7.76e−2	0.95	−1.32e−2	9.41e−2	0.95
		0.2	−1.64e−2	8.63e−2	0.94	−1.60e−2	9.45e−2	0.94
150	0.5	0.8	−5.19e−2	2.86e−2	0.94	−9.16e−3	3.39e−2	0.95
		0.5	−2.78e−2	2.44e−2	0.96	−9.79e−3	2.75e−2	0.95
		0.2	−3.68e−3	2.60e−2	0.95	−9.007e−3	2.74e−2	0.945
	0.3	0.8	−3.14e−2	2.47e−2	0.96	−9.55e−3	3.12e−2	0.955
		0.5	−7.38e−3	2.40e−2	0.96	−8.52e−3	2.77e−2	0.95
		0.2	−4.04e−3	2.58e−2	0.95	−4.38e−3	2.66e−2	0.95
	0.1	0.8	−1.12e−2	2.40e−2	0.955	−4.54e−3	3.17e−2	0.94
		0.5	−1.28e−2	2.57e−2	0.945	−2.27e−2	2.96e−2	0.94
		0.2	−9.74e−3	2.44e−2	0.96	−8.65e−3	2.67e−2	0.96

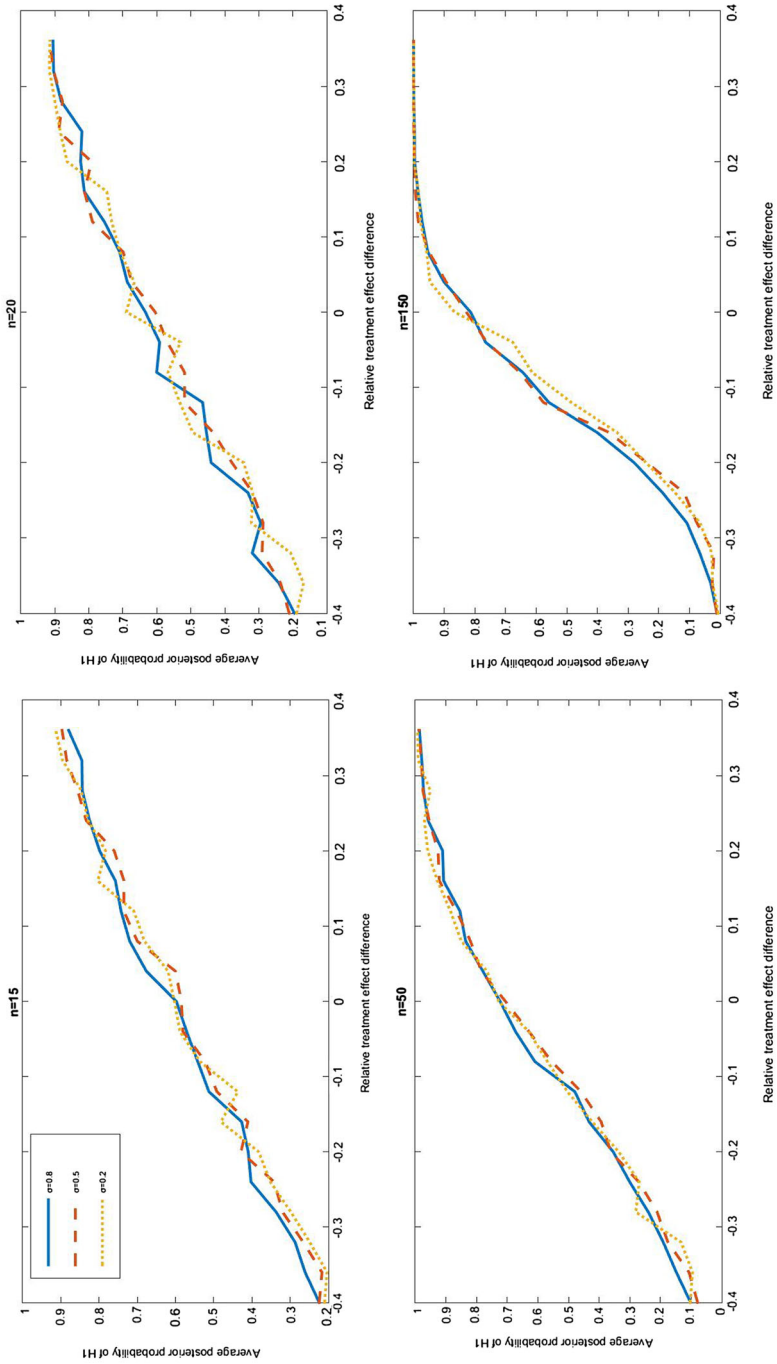


Fig. 1 Simulation: the expected posterior probability of $E[P(H_1 | \text{Data})]$ as a function of relative treatment effect difference

Table 2 Real data example 1: the frequency distribution of patients with a success (+) or a failure (–) during treatment A (200 μg salbutamol solution aerosol) or B (12 μg formoterol solution aerosol)

Sequence	(+, +)	(+, –)	(–, +)	(–, –)	Total
AB	5	1	6	0	12
BA	2	9	0	1	12

Table 3 Real data example 1: the comparison of results ν_{BA} by YB and LC under varying scenarios

Approach	YB	YB	LC	LC	LC
Data modified (each cell added a number)	0	+1	+0.5	+0.001	+1
ν_{BA}	173.33	10.72	9.074	232.30	5.92

and drug A, 200 μg salbutamol solution aerosol. The first group of 12 children received treatment A first and B second; the second group of 12 children received treatment B first and A second. The binary outcome was derived from a subjectively judged four-point scale: a success represents a good response, and a failure represents a poor, fair or moderate efficacy. Here, we want to test whether drug B is noninferior to drug A based on the odds ratio of the patient success responses rate.

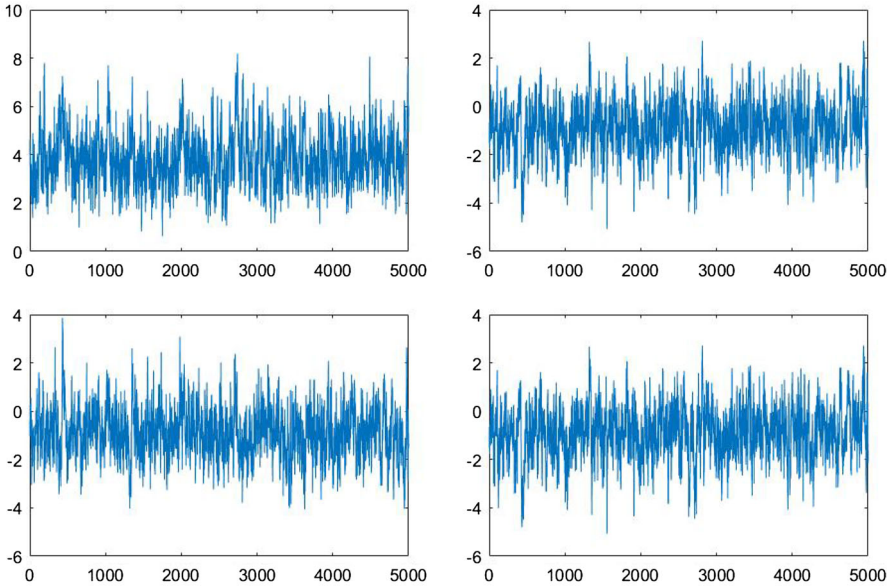
We specify the priors as in the simulation example. We run the Gibbs sampler for 5000 iterations after a 1000 burn-in period and conduct the Gelman–Rubin approach to assess convergence as in the simulation section. Similar sensitivity tests as in the simulation example are also conducted. The MCMC chains show good convergence and the results are consistent.

We obtain the mean estimate ν_{BA} and the corresponding 95% CI as 173.33 and (7.21, 628.67), respectively. We should note that this is a very small data set. In particular, there are two cells with zero counts, five cells with value no more than 2. Because of the cells with zero counts, the approach of Lui and Chang [16] can not estimate it directly. They obtain the estimate for the odds ratio as 9.074 by adding 0.5 to the cells. Given the small sample size and distribution of small values, it might still change the results significantly by adding a seemingly small value. However, what is a reasonably small value to be used is not justified by Lui and Chang [16]. For example, when we try different small values from 0.001 to 1.0, the estimated results of Lui and Chang [16] can vary from 232.30 to 5.92. We run our approach by adding 1 to the corresponding cells and obtain the estimate ν_{BA} and the corresponding 95% CI as 10.72 and (2.70, 30.22), respectively. The median value is drastically changed from 42.43 to 8.60, see Table 3. When the maximum clinically acceptable noninferior margin δ is set 0.8, we derive the posterior probability of the alternative hypothesis almost 0.0. Obviously, we can easily derive the conclusion that the treatment of 12 μg formoterol solution aerosol is noninferior to the treatment of 200 μg salbutamol solution aerosol.

For the second example, we consider a crossover trial study analyzed by Ezzet and Whitehead [15]. In the study, 3M-Riskier conducted a crossover trial to compare the suitability of two new inhalation devices (A and B) in patients who were currently using a standard inhaler device delivering Salbutamol. Patients in group 1 used device

Table 4 Real data example 2: the frequency distribution of patients with a response “yes” (+) or a “no” (–) comparing two new inhalation devices delivering salbutamol

Sequence	(+, +)	(+, –)	(–, +)	(–, –)	Total
AB	26	41	15	57	139
BA	38	16	32	54	140

**Fig. 2** Application 2: traceplots of relative treatment effect difference, group difference, period difference, and intercept

A for one week and then device B for another week. The other patients in group 2 used the devices in reverse order. No washout was felt necessary. Patients were surveyed whether there were particular features which they liked about each device, and their responses were coded as “yes” or “no.” There were less than 3% patients with missing outcomes. We assume that the missing outcomes occur completely at random and summarize the frequencies of patients with known responses in Table 4 for purpose of illustration. To assess the noninferiority of device A versus device B, we run our approach and derive the ν_{BA} and the corresponding 95% CI as 2.42 and (1.58, 3.68). When the maximum clinically acceptable noninferior margin δ is set 0.8, we derive the posterior probability of the alternative hypothesis 0.00081. All the results indicate that device A is noninferior to device B based on the odds ratio of the patient favor response. The approach of Lui and Chang [16] provides results close to ours with estimate ν 2.34 and p value less than 0.001. The traceplots of several parameters are provided in Fig. 2 to assess convergence, we can see the results converge very fast. Here, we should note that the second example has a fairly big sample size 279. For the real data examples, we show that our approach provides results very close to those of

Lui and Change [16] with fairly big data sets. However, for small data examples, our approach can provide more consistent and reliable results. We also run the analyses for the applications with varying means and variances of the priors for the parameters to evaluate the effects. For the first application with small sample size, the length of the coverage is a little wider with flatter prior, but negligible with the second application with bigger sample size. We do not see any significant deviation in the parameter estimation.

5 Conclusion and Discussion

In this paper, we have developed a Bayesian approach for hypothesis testing in non-inferiority and equivalence based on odds ratio of patient response rate for simple crossover trials. The approach can be easily implemented and improve efficiency using data augmentation and scale mixture of normal representation. Compared to the current frequentist approaches, our approach does not suffer from the asymptotic constraints and can easily handle cells with zero counts. Through simulation studies, we have shown the strength and the good performance of our approach. The approach can be easily modified to accommodate the more complicated case, for example, three-arm trials. We expect the approach can be of help for the clinical researchers conducting research in noninferiority test for simple crossover trials.

Acknowledgements The work is supported in part by the University Grant Program of San Diego State University (UGP 242546). The author thanks my colleagues for their comments and critical readings of the manuscript.

Appendix

- Given $u_0 \sim N(\mu_1, \sigma_1^2)$, the full conditional $p(u_0|\text{else})$ is sampled from a normal distribution $N(\mu_1^*, \sigma_1^{*2})$

$$\sigma_1^{*-2} = \sigma_1^{-2} + \sum_g \sum_i \sum_j \phi_{gij} / \sigma^2 \tag{11}$$

$$\mu_1^* = \sigma_1^{*2} \left\{ \mu_1 / \sigma_1^2 + \sum_g \sum_i \sum_j \phi_{gij} / \sigma^2 (y_{gij}^* - (\psi_j + t_{l(g,j)} + \mu_{g,i})) \right\} \tag{12}$$

- Given $\mu_{gi} \sim N(0, \rho^2)$, the full conditional $p(\mu_{gi}|\text{else})$ is sampled from a normal distribution $N(\mu^*, \sigma^{*2})$

$$\sigma^{*-2} = \rho^{-2} + \sum_j \phi_{gij} / \sigma^2 \tag{13}$$

$$\mu^* = \sigma^{*2} \left\{ \sum_j \phi_{gij} / \sigma^2 (y_{gij}^* - (\psi_j + t_{l(g,j)} + u_0)) \right\} \tag{14}$$

3. Given $\rho^2 \sim \text{IG}(a_0, b_0)$, the full conditional $p(\rho^2|\text{else})$ is sampled from an inverse Gamma distribution $\text{IG}(a^*, b^*)$

$$a^* = a_0 + \sum_g \sum_i 1/2 \quad (15)$$

$$b^* = b_0 + \sum_g \sum_i \mu_{gi}^2/2 \quad (16)$$

4. Given $\psi_j \sim N(\mu_3, \sigma_3^2)$, the full conditional $p(\psi_j|\text{else})$ is sampled from a normal distribution $N(\mu_3^*, \sigma_3^{*2})$

$$\sigma_3^{*-2} = \sigma_3^{-2} + \sum_g \sum_i \phi_{gij}/\sigma^2 \quad (17)$$

$$\mu_3^* = \sigma_3^{*2} \left\{ \mu_3/\sigma_3^2 + \sum_g \sum_i \phi_{gij}/\sigma^2 (y_{gij}^* - (\mu_{gi} + t_{l(g,j)} + u_0)) \right\} \quad (18)$$

5. Given $t_m \sim N(\mu_4, \sigma_4^2)$, the full conditional $p(t_m|\text{else})$ is sampled from a normal distribution $N(\mu_4^*, \sigma_4^{*2})$

$$\sigma_4^{*-2} = \sigma_4^{-2} + \sum_g \sum_j \sum_i I(l(g, j) = m) \phi_{gij}/\sigma^2 \quad (19)$$

$$\mu_4^* = \sigma_4^{*2} \left\{ \mu_4/\sigma_4^2 + \sum_g \sum_i \sum_j I(l(g, j) = m) \phi_{gij}/\sigma^2 \right. \\ \left. \times (y_{gij}^* - (\mu_{gi} + \psi_j + u_0)) \right\} \quad (20)$$

6. Given $\phi_{gij} \sim \text{IG}(v/2, v/2)$, the full conditional $p(\phi_{gij}|\text{else})$ is sampled from an inverse Gamma distribution $\text{IG}(a_1^*, b_1^*)$

$$a_1^* = (v + 1)/2 \quad (21)$$

$$b_1^* = v/2 + (y_{gij}^* - (\psi_j + \mu_{gi} + t_{l(g,j)} + u_0))/(2\sigma^2) \quad (22)$$

References

- Chen J, Tsong Y, Kang S (2000) Tests for equivalence or noninferiority between two proportions. *Drug Inf J* 34:569–578
- Liu J, Chow S (1993) On assessment of bioequivalence of drugs with negligible plasma levels. *Biom J* 35:109–123
- Chow S, Liu J (2009) Design and analysis of bioavailability and bioequivalence studies. Taylor & Francis, New York
- Tu D (1998) On the use of the ratio or the odds ratio of cure rates in establishing therapeutic equivalence of non-systemic drugs with binary clinical endpoints. *J Biopharm Stat* 8:263–282
- Tu D (2003) Encyclopedia of biopharmaceutical statistics. Dekker, New York

6. Tu D (2003) Statistical controversies in design and analysis of bioequivalence trial for pharmaceuticals with negligible blood levels: the metered dose intraler trial. In: The 14th midwest biopharmaceutical statistics workshop, Muncie, Indiana
7. Lui K, Chang K (2015) Test and estimation in binary data analysis under an incomplete block crossover design. *Comput Stat Data Anal* 81:130–138
8. Herson J (1991) *Encyclopedia of biopharmaceutical statistics*. Dekker, New York
9. EMEA European Medicines Agency (2005) *Guideline on the choice of the noninferiority margin*
10. Fleiss J, Levin B, Park M (2003) *Statistical methods for rates and proportions*, 3rd edn. Wiley, New York
11. Fleiss J (1986) *The design and analysis of clinical experiments*. Wiley, New York
12. Gart J (1969) An exact test for comparing matched proportions in crossover design. *Biometrika* 56:75–80
13. Zimmermann H, Rahlfs V (1978) Test hypotheses in the two period change over with binary data. *Biom J* 20:133–141
14. Schouten H, Kester A (2010) A simple analysis of a simple crossover trial with a dichotomous outcome measure. *Stat Med* 29:193–198
15. Ezzet F, Whitehead J (1992) A random effects model for binary data from crossover clinical trials. *Appl Stat* 41:117–126
16. Lui K, Chang K (2011) Test noninferiority (and equivalence) based on the odds ratio under a simple crossover trial. *Stat Med* 11:1230–1242
17. Senn S (2002) *Cross-over trials in clinical research*, 2nd edn. Wiley, Chichester
18. Senn S (2006) Cross-over trials in statistics in medicine: the first 25 years. *Stat Med* 25:3430–3442
19. Osman M, Ghosh S (2011) Semiparametric Bayesian testing procedure for noninferiority trials with binary endpoints. *J Biopharm Stat* 21:920–937
20. Ghosh P, Nathoo F, Gonen M, Tiwari R (2010) Assessing noninferiority in a three-arm trial using the Bayesian approach. *Stat Med* 15:1795–1808
21. Basu S, Santra S (2010) A joint model for incomplete data in crossover trials. *J Stat Plan Inference* 140:2839–2845
22. Grieve A (1985) A Bayesian analysis of the two-period crossover design for clinical trials. *Biometrics* 41:979–990
23. Albert J, Chib S (1997) Bayesian tests and model diagnostics in conditionally independent hierarchical models. *J Am Stat Assoc* 92:916–925
24. Holmes C, Knorr-Held L (2003) *Efficient simulation of Bayesian logistic regression models*. Technical report, Ludwig Maximilians University Munich
25. O'Brien SM, Dunson DB (2004) Bayesian multivariate logistic regression. *Biometrics* 60:739–746
26. West M (1987) On scale mixtures of normal distributions. *Biometrika* 74:646–648
27. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–511