Assessing non-inferiority for incomplete paired-data under non-ignorable missing mechanism

Huiqiong Li\textsuperscript{a}, Guoliang Tian\textsuperscript{b}, Niansheng Tang\textsuperscript{a}, Hongyuan Cao\textsuperscript{c,d,*}

\textsuperscript{a} Department of Statistics, Yunnan University, PR China
\textsuperscript{b} Department of Mathematics, Southern University of Science and Technology, PR China
\textsuperscript{c} Center for Applied Statistical Research, School of Mathematics, Jilin University, PR China
\textsuperscript{d} Department of Statistics, University of Missouri-Columbia, USA

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 20 December 2016
Received in revised form 29 April 2018
Accepted 12 May 2018
Available online 26 May 2018

\textbf{Keywords:}
Bayesian p-values
Credible intervals
Highest posterior density intervals
Non-inferiority test
Incomplete paired data

\textbf{ABSTRACT}

Testing equivalence of incomplete paired data arises frequently in biomedical studies. Most existing work impose the missing at random assumption, which is not realistic in practice. Two Bayesian approaches for testing the non-inferiority of incomplete paired data under non-ignorable missing mechanism are presented. In addition, Bayesian credible intervals and highest posterior density intervals for the risk difference are constructed. Simulation studies are conducted to evaluate the performance of the two Bayesian testing procedures and the credible intervals. Two datasets are used to illustrate the proposed methods.

\textcopyright 2018 Elsevier B.V. All rights reserved.

1. Introduction

Assessing the non-inferiority of a new method or treatment with the standard one is an important topic in comparative clinical studies. Non-inferiority trials are often employed to evaluate whether a less toxic, easier to administer or inexpensive new treatment is not inferior to the standard treatment in terms of efficacy. Non-inferiority assessment has received a lot of attention for matched-pair trials in the past decades. For example, Tango (1998) derived a score statistic to test non-inferiority via relative risk in a re-parameterized model with a matched-pair design. Tang et al. (2003) developed an alternative score test procedure to test equivalence or non-inferiority via relative risk in a matched-pair design. Chan et al. (2003) proposed an exact method to assess non-inferiority via rate ratio with small-sample matched-pair design.

In practice, in comparative studies of two treatments or reviewers, incomplete matched-pair data are often encountered. For example, in a study of medical malpractice cases (Greenberg et al., 2007; Lin et al., 2009; Altham and Hankin, 2010; Konietschke et al., 2012), two surgeon-reviewers used a structured instrument to evaluate 69 errors, and to identify important human and system factors contributing to the errors. Among many possible factors is communication breakdown, each surgeon-reviewer was asked to determine whether a handoff in care was associated with the communication breakdown. In this study, 8 reviews were missing for Surgeon 1 and 11 reviews were missing for Surgeon 2. Thus, the resultant data include two parts: the complete observations and the incomplete observations. This dataset is displayed in Table 1.

Under the assumption of missing at random (MAR), the probability of missing only depends on observed data. In the case of MAR, various authors have studied the problem of the equivalence test and confidence interval construction for
two correlated proportions with incomplete matched-pair data (Choi and Stablein, 1982; Ekbohm, 1982; Tang and Tang, 2004; Tang et al., 2009; Lin et al., 2009; Tang et al., 2011). Non-ignorable missing or missing not at random (MNAR) refers to the case that the probability of missing is related to the value of the missing data. In the case of MNAR, Choi and Stablein (1988) proposed several methods for testing the equality of two correlated proportions. Nandram and Choi (2002) proposed a Bayesian approach for a non-ignorable non-response model. To the best of our knowledge, there is no published work to date that deals with incomplete paired-data under non-ignorable missing mechanism. In this paper, we develop Bayesian methods to test non-inferiority and to construct Bayesian credible intervals and highest posterior density (HPD) intervals for incomplete paired-data.

The rest of this paper is organized as follows. In Section 2, we present two Bayesian p-values to assess non-inferiority for incomplete paired data under the non-ignorable missing mechanism. Section 3 develops a new Bayesian interval estimation of the risk difference for incomplete paired data under nonignorable missing mechanism. Simulation studies are conducted to investigate the performance of various methods in Section 4. We illustrate the proposed methodology with two dataset in Section 5. Concluding remarks are given in Section 6.

2. Bayesian methods for the non-inferiority test

2.1. Data structure and non-inferiority test

Consider a trial for comparing two treatments. Suppose that \( X \) and \( Y \) are two correlated binary variables. Let \( X = 1 \) (or \( X = 0 \)) if a subject has a positive (or negative) response under treatment 1 and let \( Y = 1 \) (or \( Y = 0 \)) if the same subject has a positive (or negative) response under treatment 2. Let \( \theta = (\theta_1, \ldots, \theta_4)^T \) denote model parameters, where \( \theta_1 = \Pr(X = 1, Y = 1) \), \( \theta_2 = \Pr(X = 1, Y = 0) \), \( \theta_3 = \Pr(X = 0, Y = 1) \) and \( \theta_4 = \Pr(X = 0, Y = 0) \). Naturally, we have \( \Theta \subseteq \mathbb{R}_+^4 \), where \( \Theta := \{ (x_1, \ldots, x_4)^T : x_i > 0, \sum_{i=1}^4 x_i = 1 \} \).

Suppose that in a comparative trial there are a total of \( N \) participants containing \( n = \sum_{j=1}^4 n_j \) complete cases and \( m_{12} + m_{34} + m_{13} + m_{24} + m_{1234} \) incomplete cases, where \( n_1 \) subjects have both positive responses, \( n_2 \) subjects have a positive response for Treatment 1 and a negative response for Treatment 2, \( n_3 \) subjects have a negative response for Treatment 1 and a positive response for Treatment 2, \( n_4 \) subjects have both negative responses; \( m_{12} \) (or \( m_{34} \)) subjects only have a positive (or negative) response for Treatment 1, \( m_{13} \) (or \( m_{24} \)) subjects only have a positive (or negative) response for Treatment 2; and the responses for \( m_{1234} \) subjects are totally missing for both treatments. These observed outcomes are reported in Table 2. We denote the observed data by \( \mathbf{Y}_{\text{obs}} = \{ n_1, \ldots, n_4; m_{12}, m_{34}, m_{13}, m_{24}, m_{1234} \} \) with \( N = \sum_{j=1}^n n_j + m_{12} + m_{34} + m_{13} + m_{24} + m_{1234} \).

Treatment 1 is said to be not inferior to Treatment 2 if \( \Pr(X = 1) > \Pr(Y = 1) - \delta_0 \), i.e., \( \theta_1 + \theta_2 > \theta_1 + \theta_3 - \delta_0 \), where \( \delta_0 > 0 \) is the non-inferiority margin of clinical interest. Thus, testing the non-inferiority of Treatment 1 to Treatment 2 is equivalent to testing the following hypothesis:

\[
H_0: \theta_2 \leq \theta_3 - \delta_0 \quad \text{against} \quad H_1: \theta_2 > \theta_3 - \delta_0. \tag{2.1}
\]

The objective of this paper is to develop Bayesian methods for testing \( H_0 \) versus \( H_1 \) under the non-ignorable missing mechanism.

2.2. Formulation of the non-ignorable missing mechanism

To describe the non-ignorable missing mechanism in Table 2, we first define a 4-category response random variable \( R \), where \( R = 12 \) if a subject has response to both treatments, \( R = 12 \) if a subject has response only to Treatment 1, \( R = 12 \)
Table 3
The observed counts, the marginal probabilities of Z and R, and the joint distribution of \((Z, R)\).

<table>
<thead>
<tr>
<th>((Z, R))</th>
<th>Response variable (R)</th>
<th>Marginal probability of (Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1, 1))</td>
<td>(\pi_{11} n_1)</td>
<td>(\phi_1)</td>
</tr>
<tr>
<td>((1, 0))</td>
<td>(\pi_{12} n_2)</td>
<td>(\phi_2)</td>
</tr>
<tr>
<td>((0, 1))</td>
<td>(\pi_{13} n_3)</td>
<td>(\phi_3)</td>
</tr>
<tr>
<td>((0, 0))</td>
<td>(\pi_{14} n_4)</td>
<td>(\phi_4)</td>
</tr>
</tbody>
</table>

Marginal probability of \(R\)

\(\phi_1 \quad \phi_2 \quad \phi_3 \quad \phi_4 \quad 1\)

Observed counts

\(n \quad m_{12} \quad m_{34} \quad m_{13} \quad m_{24} \quad m_{1234} \quad N\)

if a subject has response only to Treatment 2 and \(R = 1\), if a subject has no response to both treatments. The probability mass function of \(R\) is represented by \(\phi_1 = \Pr(R = 12), \phi_2 = \Pr(R = 12), \phi_3 = \Pr(R = 12)\), and \(\phi_4 = \Pr(R = 12)\). Let \(\phi = (\phi_1, \ldots, \phi_4)^T\). Then \(\phi\) is the parameter vector associated with the missing data mechanism.

For convenience, define \(Z = (X, Y)\). The joint distribution of \(Z\) and \(R\) is denoted by \(\pi = (\pi_{ij})\) for \(i, j = 1, \ldots, 4\). For example, \(\pi_{11} = \Pr[Z = (1, 1), R = 12]\) and \(\pi_{44} = \Pr[Z = (0, 0), R = 12]\). Table 3 shows the observed counts, the marginal probabilities of \(Z\) and \(R\), and the joint distribution of \((Z, R)\). The likelihood function for \(\pi\) based on the observed data \(D_{\text{obs}}\) is proportional to (Tian et al., 2003)

\[
\prod_{j=1}^{4} \pi_{ij}^{n_{ij}} (\prod_{j=1}^{4} \theta_j^{n_{ij}}) m_{12}^{m_{12}} m_{34}^{m_{34}} \theta_1^{m_{12}} \theta_2^{m_{34}} \theta_3^{m_{13}} \theta_4^{m_{24}} \tag{2.2}
\]

If \(Z\) and \(R\) are independent, \(\pi_{ij} = \phi_1 \theta_{ij}, \ i, j = 1, \ldots, 4\),

the missing data mechanism is ignorable or MAR (Little and Rubin, 2002). As \(\sum_{i=1}^{4} \theta_i = 1\) and \(\sum_{i=1}^{4} \phi_i = 1\), under MAR, there are only 6 free parameters and (2.2) becomes

\[
\prod_{j=1}^{4} \theta_j^{n_{ij}} \prod_{j=1}^{4} \phi_i^{n_{ij}} \theta_1^{m_{12}} \theta_2^{m_{34}} \theta_3^{m_{13}} \theta_4^{m_{24}} \times \phi_1^{\theta_1} \phi_2^{\theta_2} \phi_3^{\theta_3} \phi_4^{\theta_4} \tag{2.3}
\]

Based on (2.3), we can obtain the maximum likelihood estimates (MLEs) of \(\theta\) and \(\phi\) with explicit expressions. In other words, both \(\theta\) and \(\phi\) are estimable in the frequentist framework.

If \(Z\) and \(R\) are dependent, we call the missing data mechanism non-ignorable. Under the non-ignorable missing mechanism, we have 15 free parameters since \(\pi \in \mathbb{T}_{16}\). These \(\pi_{ij}\) are not estimable in the frequentist framework. Yet, they are identifiable if we put a prior distribution on \(\pi\).

2.3. Bayesian tests

The Bayes factor is a very useful tool to test hypothesis. The Bayes factor for comparing \(H_1\) to \(H_0\) is given by

\[
BF_{10} = \frac{m_1(D_{\text{obs}})}{m_0(D_{\text{obs}})} \tag{2.4}
\]

where

\[
m_j(D_{\text{obs}}) = \int \int p(D_{\text{obs}}|\pi_j) f(\pi_j) d\pi_j, \tag{2.5}
\]

where \(p(D_{\text{obs}}|\pi_j)\) is the likelihood function of the observed data and \(f(\pi_j)\) is the prior for the unknown parameter under \(H_j, j = 0, 1\), respectively.

The Bayes factor provides evidence in favor of \(H_1\) against \(H_0\). According to Kass and Raftery (1995), if \(0 \leq \log(BF_{10}) < 1\), the evidence against \(H_0\) is “not worth more than a bare mention”; if \(1 \leq \log(BF_{10}) < 3\), the evidence against \(H_0\) is “positive”; if \(3 \leq \log(BF_{10}) < 5\), the evidence against \(H_0\) is “strong”; and if \(\log(BF_{10}) \geq 5\), the evidence against \(H_0\) is “very strong”.

However, the function \(m_j(D_{\text{obs}})\) in (2.5) usually does not have a closed form and the integration is extremely difficult to calculate. To circumvent this, we shall develop Bayesian \(p\)-values to test \(H_0\) against \(H_1\) in (2.1) for incomplete paired-data under non-ignorable missing mechanism.
2.3.1. The conventional Bayesian \( p \)-value

In order to test (2.1) under non-ignorable missing mechanism, we first discuss the Bayesian \( p \)-value with complete data. Suppose the joint prior density of \((\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14})\) is Dirichlet, with parameters \(\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14})\), given by the density function
\[
f(\pi) \propto \pi_{11}^{\alpha_{11}-1} \pi_{12}^{\alpha_{12}-1} \pi_{13}^{\alpha_{13}-1} \pi_{14}^{\alpha_{14}-1}.
\]
Then, if there are no missing data, the joint posterior distribution of \((\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14})\) is proportional to
\[
\prod_{j=1}^{4} \pi_{1j}^{\eta_{j} + \alpha_{1j} - 1}.
\]
Based on this posterior distribution, \( p \)-value can be obtained for the one-sided alternative \(\theta_{2} > \theta_{3} - \delta_{0}\) using the tail probability (Altham, 1969, 1971)
\[
p = \Pr(\theta_{2} \leq \theta_{3} - \delta_{0}|D_{\text{obs}}).
\]
For the improper Dirichlet prior with \(\alpha_{12} = 0, \alpha_{13} = 1 \text{ and } \alpha_{11} = \alpha_{14} = 0\), Altham (1971) showed the posterior probability that \(\theta_{2} \leq \theta_{3} - \delta_{0}\) is identical to the exact McNemar \( p \)-value (Mosteller, 1952) for the one-sided alternative \(\theta_{2} > \theta_{3} - \delta_{0}\).

Using the same Dirichlet prior, we extend the approach of Altham (1969, 1971) to develop Bayesian \( p \)-values under non-ignorable missing mechanism. By calculating the posterior probability, the proposed Bayesian \( p \)-value for the one-sided alternative \(H_{1} : \theta_{2} > \theta_{3} - \delta_{0}\) is the tail probability
\[
p_{e1} = \Pr(\theta_{2} \leq \theta_{3} - \delta_{0}|D_{\text{obs}}).
\] (2.6)
To compute the posterior probability under non-ignorable missing mechanism, one need to determine priors for \((\pi_{ij})\). Let the joint prior for \((\pi_{ij})\) be a Dirichlet distribution with known parameters \((\alpha_{ij})\); that is \(f(\pi) \propto \prod_{i=1}^{4} \prod_{j=1}^{4} \pi_{ij}^{\alpha_{ij} - 1}\). Then, the posterior distribution by combining the prior distribution and the likelihood function can be calculated. Through (2.2), the posterior distribution of \(\pi\) is proportional to
\[
\left(\prod_{j=1}^{4} \pi_{1j}^{\eta_{j} + \alpha_{1j} - 1}\right) \left(\prod_{i=2}^{4} \prod_{j=1}^{4} \pi_{ij}^{\alpha_{ij} - 1}\right) (\pi_{21} + \pi_{22})^{m_{12}} (\pi_{23} + \pi_{24})^{m_{13}} \prod_{j=1}^{4} \pi_{4j}^{m_{1234}}
\times (\pi_{31} + \pi_{33})^{m_{13}} (\pi_{32} + \pi_{34})^{m_{124}} \prod_{j=1}^{4} \pi_{4j}^{m_{1234}} .
\] (2.7)
Unfortunately, under non-ignorable missing mechanism, for any Dirichlet prior, the Bayesian \( p \)-value in (2.6) does not have a closed form. Since the dimension of the integration is high, (2.6) can be extremely difficult to calculate by direct numerical integration. Here, we shall compute it by sampling-based method.
First, we have the following notations.
\[
\begin{align*}
\pi_{1i} & = \xi_{i}, \quad 1 \leq i \leq 4, \\
\pi_{2,i-4} & = \xi_{i}, \quad 5 \leq i \leq 8, \\
\pi_{31} & = \xi_{9}, \\
\pi_{32} & = \xi_{10}, \\
\pi_{33} & = \xi_{11}, \\
\pi_{34} & = \xi_{12}, \\
\pi_{4,i-12} & = \xi_{i}, \quad 13 \leq i \leq 16.
\end{align*}
\] (2.8)
It can be shown that the posterior distribution in (2.7) becomes
\[
\left(\prod_{i=1}^{4} \xi_{i}^{\eta_{i} + \alpha_{1i} - 1}\right) \left(\prod_{i=5}^{8} \xi_{i}^{\alpha_{2,i-4} - 1}\right) \xi_{9}^{\alpha_{31} - 1} \xi_{10}^{\alpha_{32} - 1} \xi_{11}^{\alpha_{33} - 1} \xi_{12}^{\alpha_{34} - 1} \left(\prod_{i=13}^{16} \xi_{i}^{\alpha_{4,i-12} - 1}\right) \\
\times \left(\sum_{j=1}^{4} \xi_{j}\right)^{m_{12}} \left(\sum_{j=5}^{8} \xi_{j}\right)^{m_{13}} \left(\sum_{j=9}^{12} \xi_{j}\right)^{m_{124}} \left(\sum_{j=13}^{16} \xi_{j}\right)^{m_{1234}} .
\] (2.9)
According to Tian et al. (2003), (2.9) is the kernel of a grouped Dirichlet distribution (GDD) with six partitions. The definition of GDD is given as follows.
An \( n \)-dimensional random vector \( \mathbf{X} = (X_1, \ldots, X_n)\) is said to follow a grouped Dirichlet distribution with \( m \) partitions if the joint distribution of \((X_1, \ldots, X_{n-1})\) is given by

\[
\frac{1}{c_m} \left( \prod_{i=1}^n x_i^{a_i-1} \right) \cdot \prod_{j=1}^m \left( \sum_{k=j-1+1}^{s_j} x_k \right)^{b_j},
\]

where \( c_m \) is the normalizing constant given by

\[
c_m = \left\{ \prod_{j=1}^m B_{j-s_j-1}(a_{j-1+1}, \ldots, a_j) \right\} \cdot B_m \left( \sum_{k=1}^{s_1} a_k + b_1, \ldots, \sum_{k=s_{m-1}+1} a_k + b_m \right)
\]

and \( B_n(a_1, \ldots, a_n) = \left\{ \prod_{j=1}^n f'(a_j) / f'(\sum_{j=1}^n a_j) \right\} \) denotes the multivariate beta function. Write \( \mathbf{X} \sim \text{GD}_n, m, \mathbf{a}, \mathbf{b} \) on \( \mathbb{T}_n \), where \( \mathbf{a} = (a_1, \ldots, a_n)^T \) is a positive vector, \( \mathbf{b} = (b_1, \ldots, b_m)^T \) is a non-negative vector, and \( \mathbf{s} = (s_1, \ldots, s_m)^T \) with \( 0 := s_0 < 1 < s_1 < \cdots < s_m := n \).

According to Tian et al. (2003), if an \( n \)-dimensional random vector \( \mathbf{X} \sim \text{GD}_n, m, \mathbf{a}, \mathbf{b} \) on \( \mathbb{T}_n \), we can partition \( \mathbf{X} \) into block vector \((\mathbf{X}_1, \ldots, \mathbf{X}_m)\) with the following stochastic representation:

\[
\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\
\vdots \\
\mathbf{X}_m \end{pmatrix} \overset{d}{=} \begin{pmatrix} \mathbf{R}_1 \cdot \mathbf{Y}_1 \\
\vdots \\
\mathbf{R}_m \cdot \mathbf{Y}_m \end{pmatrix},
\]

(2.10)

where (i) \( \mathbf{R} = (R_1, \ldots, R_m)^T \sim \text{Dirichlet}(\beta_1, \ldots, \beta_m) \) with \( \beta_j = \sum_{k=j-1+1}^{s_j} a_k + b_j, j = 1, \ldots, m; \) (ii) \( \mathbf{Y}_j \sim \text{Dirichlet}(\mathbf{a}_j), j = 1, \ldots, m; \) and (iii) \( \mathbf{R} \) and \((\mathbf{Y}_1, \ldots, \mathbf{Y}_m)\) are mutually independent. The notation \( \mathbf{X} \overset{d}{=} \mathbf{Y} \) means that the random vectors \( \mathbf{X} \) and \( \mathbf{Y} \) have the same distribution.

Let \( \xi = (\xi_1, \ldots, \xi_{10})^T \), we have \( \xi | \mathbf{D}_{\text{obs}} \sim \text{GD}_{16, 6, 4}(\mathbf{a}, \mathbf{b}) \), where \( \mathbf{a} = (n_1 + \alpha_{11}, n_2 + \alpha_{12}, n_3 + \alpha_{13}, \ldots, n_{10} + \alpha_{10}, \alpha_{21}, \ldots, \alpha_{24}, \alpha_{31}, \alpha_{32}, \alpha_{33}, \alpha_{34}, \alpha_{41}, \alpha_{42}, \alpha_{43}, \alpha_{44})^T \), \( \mathbf{b} = (0, m_{12}, m_{34}, m_{13}, m_{24}, m_{1234})^T \) and \( \mathbf{s} = (4, 6, 8, 10, 12, 16)^T \). Thus, we can generate i.i.d. samples from (2.10).

Next, from Table 3, note that \( \theta_2, \theta_3 \) and \( \xi \) have the following relationship:

\[
\theta_2 = \sum_{i=1}^4 \pi_{i2} = \xi_2 + \xi_6 + \xi_{11} + \xi_{14} \quad \text{and} \quad \theta_3 = \sum_{i=1}^4 \pi_{i3} = \xi_3 + \xi_7 + \xi_{10} + \xi_{15}.
\]

Thus, the Bayesian \( p \)-value can be computed from the following sampling-based method.

**Step 1.** Generate i.i.d. posterior sample \( \xi^{(1)}, \ldots, \xi^{(G)} \) from \( \text{GD}_{16, 6, 4}(\mathbf{a}, \mathbf{b}) \), where \( \xi^{(g)} = (\xi_1^{(g)}, \ldots, \xi_{16}^{(g)})^T, g = 1, \ldots, G \).

**Step 2.** Based on \( \xi^{(g)} \), calculate \( \theta_2^{(g)} = \xi_2^{(g)} + \xi_6^{(g)} + \xi_{11}^{(g)} + \xi_{14}^{(g)} \) and \( \theta_3^{(g)} = \xi_3^{(g)} + \xi_7^{(g)} + \xi_{10}^{(g)} + \xi_{15}^{(g)} \) for \( g = 1, \ldots, G \).

**Step 3.** The Bayesian \( p \)-value is given by

\[
p_{\text{v1}} = \frac{1}{G} \sum_{g=1}^G I(\theta_2^{(g)} \leq \theta_3^{(g)} - \delta_0),
\]

where \( I(\cdot) \) is the indicator function.

2.3.2. The adjusted Bayesian \( p \)-value

Lindley (1957) pointed out a paradoxical result that even when a very small observed significance level is obtained, the posterior probability of being null can be close to one as long as the sample size is large enough. This observation motivated us to provide an alternative Bayesian test based on the method proposed in Yin (2012).

Specifically, let \( \Delta = \theta_2 - \theta_3 - \delta_0 \), the hypothesis in (2.1) becomes \( H_0: \Delta \leq 0 \) against \( H_1: \Delta > 0 \). Hence, a Bayesian measure for testing \( H_0 \) against \( H_1 \) can be defined by

\[
p_{\text{v2}} = \Pr\left\{ \Delta - E(\Delta|\mathbf{D}_{\text{obs}}) \leq -|E(\Delta|\mathbf{D}_{\text{obs}})| \middle| \mathbf{D}_{\text{obs}} \right\},
\]

(2.11)

where \( E(\Delta|\mathbf{D}_{\text{obs}}) \) is the posterior expectation of \( \Delta \), which is given by

\[
E(\Delta|\mathbf{D}_{\text{obs}}) = E(\xi_2|\mathbf{D}_{\text{obs}}) + E(\xi_6|\mathbf{D}_{\text{obs}}) + E(\xi_{11}|\mathbf{D}_{\text{obs}}) + E(\xi_{14}|\mathbf{D}_{\text{obs}})
- E(\xi_3|\mathbf{D}_{\text{obs}}) - E(\xi_7|\mathbf{D}_{\text{obs}}) - E(\xi_{10}|\mathbf{D}_{\text{obs}}) - E(\xi_{15}|\mathbf{D}_{\text{obs}}) - \delta_0.
\]

(2.12)
Ng et al. (2008) showed that the expectation of any component of a random vector following a GDD has an explicit expression. From (6.5) in Ng et al. (2008), we obtain

\[
E(\xi_j|D_{\text{obs}}) = \begin{cases} \frac{n_2 + \alpha_{12}}{N + \alpha}, \\ \frac{\alpha_{22}(\alpha_{21} + \alpha_{22} + m_{12})}{(\alpha_{21} + \alpha_{22})(N + \alpha)}, \\ \frac{\alpha_{32}(\alpha_{32} + \alpha_{34} + m_{24})}{(\alpha_{32} + \alpha_{34})(N + \alpha)}, \\ \frac{\alpha_{42}(\alpha_4 + m_{1234})}{\alpha_4(N + \alpha)}, \\ \frac{\alpha_{43}(\alpha_4 + m_{1234})}{\alpha_4(N + \alpha)}, \\ \frac{n_3 + \alpha_{13}}{N + \alpha}, \\ \frac{\alpha_{23}(\alpha_{23} + \alpha_{24} + m_{24})}{(\alpha_{23} + \alpha_{24})(N + \alpha)}, \\ \frac{\alpha_{33}(\alpha_{31} + \alpha_{33} + m_{13})}{(\alpha_{31} + \alpha_{33})(N + \alpha)}, \\ \frac{\alpha_{43}(\alpha_4 + m_{1234})}{\alpha_4(N + \alpha)}, \\ \frac{\alpha_{44}(\alpha_4 + \alpha_{12} + \alpha_{13} + \alpha_{24} + m_{24} + m_{12} + m_{13} + m_{24} + m_{1234})}{N + \alpha}. \end{cases}
\] (2.13)

where \( N = \sum_{i=1}^{4} n_i + m_{12} + m_{34} + m_{13} + m_{24} + m_{1234}. \) \( \alpha = \sum_{j=1}^{4} \alpha_j \) and \( \alpha_\cdot = \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_{ij}. \) In other words, we can compute \( E(\Delta|D_{\text{obs}}) \) exactly.

We propose the following sampling-based method for calculating the posterior probability \( p_{v2}: \)

Step 1. Generate i.i.d. samples \( \xi^{(1)}, \ldots, \xi^{(G)} \) from GD\( (\alpha, \bf{b}) \), where \( \xi^{(g)} = (\xi_1^{(g)}, \ldots, \xi_{16}^{(g)})^T, g = 1, \ldots, G. \)

Step 2. Based on \( \{\xi^{(g)}\}_{g=1}^{G} \), calculate \( \theta_2^{(g)} = \xi_2^{(g)} + \xi_6^{(g)} + \xi_{11}^{(g)} + \xi_{14}^{(g)} \) and \( \theta_3^{(g)} = \xi_3^{(g)} + \xi_7^{(g)} + \xi_{10}^{(g)} + \xi_{15}^{(g)} \) for \( g = 1, \ldots, G. \)

Step 3. Compute the posterior expectation \( E(\Delta|D_{\text{obs}}) \) through (2.12) and (2.13).

Step 4. The Bayesian \( p \)-value in (2.11) can be obtained through

\[
p_{v2} = \frac{1}{G} \sum_{g=1}^{G} I\left( \theta_2^{(g)} - \theta_3^{(g)} - \delta_0 - E(\Delta|D_{\text{obs}}) \leq -E(\Delta|D_{\text{obs}}) \right).
\]

3. Bayesian credible intervals for risk difference

The construction of confidence intervals for parameters of interest in matched-pair design has been extensively explored in the literature (Liu et al., 2002; Newcombe, 2003). In the Bayesian framework, for two binary outcomes in a \( 2 \times 2 \) contingency table, Agresti and Min (2005) and Hashemi et al. (1998) examined Bayesian credible intervals for parameters such as proportion difference, relative risk and odds ratio. For matched-pair designs with incomplete observations under MAR, Shi and Bai (2008, 2009) explored Bayesian credible intervals. For incomplete paired-data under non-ignorable missing mechanism, Bayesian credible intervals are less explored in the literature. In this section, we shall consider the construction of two Bayesian credible intervals (i.e., equal-tail credible interval and HPD interval) for risk difference in matched-pair designs with incomplete observations under non-ignorable missing mechanism.

In Section 2.3.1, we have generated i.i.d. posterior samples \( \{\theta_2^{(g)}, \theta_3^{(g)}\}_{g=1}^{G} \), which can be used to calculate the risk difference \( \delta = \theta_2 - \theta_3. \) Consequently, the \((1 - \alpha)\) percentage equal-tail credible interval for \( \delta \) is given by

\[
[\hat{\delta}_L, \hat{\delta}_U],
\] (3.1)

where \( \hat{\delta}_L \) and \( \hat{\delta}_U \) are the 100(\( \alpha \)/2) and 100(1 - \( \alpha \)/2) percentiles of \( \{\delta^{(g)}\}_{g=1}^{G} \), where \( \delta^{(g)} = \theta_2^{(g)} - \theta_3^{(g)}. \)

To find the \((1 - \alpha)\) 100\% highest posterior density (HPD) interval for \( \delta \), we first sort the i.i.d. posterior samples \( \{\delta^{(g)}\}_{g=1}^{G} \) to obtain their order statistics, denoted by \( \{\delta^{(g)}\}_{g=1}^{G}. \) The \((1 - \alpha)\) 100\% HPD interval for \( \delta \) can be obtained by minimizing the width of the following \((1 - \alpha)\) 100\% Bayesian credible intervals of \( \delta: \)

\[
[\delta^{(g)}, \delta_{\lceil(G(1-\alpha)/2)\rceil}] , g = 1, \ldots, \lceil G \alpha \rceil,
\] (3.2)

where \( \lceil r \rceil \) is the largest integer not greater than \( r \).

To demonstrate, let \( G = 1000 \) and \( \alpha = 0.05. \) We can then have 50 Bayesian credible intervals for \( \delta \) with 95\% coverage, i.e., \( \{\delta^{(g)}, \delta_{\lceil 950 + (1-\alpha)g \rceil}\}, g = 1, \ldots, 50. \) The 95\% HPD interval for \( \delta \) is \( \{\delta^{(g)}, \delta_{\lceil 950 + (1-\alpha)g \rceil}\}, \) whose width \( \delta_{\lceil 950 + (1-\alpha)g \rceil} - \delta^{(g)} \) is the shortest among the 50 Bayesian credible intervals. In fact, the 95\% equal-tail credible interval for \( \delta \) is given by \( \{\delta^{(25)}, \delta^{(975)}\}. \)
Table 4
Empirical type I error based on 10,000 replications and 5000 sampling-based samples.

<table>
<thead>
<tr>
<th>N</th>
<th>π</th>
<th>α = 0.5 \cdot 1_{16}</th>
<th>α = 1.0 \cdot 1_{16}</th>
<th>α = 1.5 \cdot 1_{16}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p_{v1}</td>
<td>p_{v2}</td>
<td>p_{v1}</td>
</tr>
<tr>
<td>30</td>
<td>Case a</td>
<td>0.034</td>
<td>0.109</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Case b</td>
<td>0.024</td>
<td>0.057</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Case c</td>
<td>0.011</td>
<td>0.043</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Case d</td>
<td>0.012</td>
<td>0.040</td>
<td>0.01</td>
</tr>
<tr>
<td>50</td>
<td>Case a</td>
<td>0.028</td>
<td>0.105</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Case b</td>
<td>0.023</td>
<td>0.062</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>Case c</td>
<td>0.007</td>
<td>0.036</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Case d</td>
<td>0.01</td>
<td>0.035</td>
<td>0.009</td>
</tr>
<tr>
<td>100</td>
<td>Case a</td>
<td>0.039</td>
<td>0.137</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Case b</td>
<td>0.025</td>
<td>0.053</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Case c</td>
<td>0.005</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Case d</td>
<td>0.009</td>
<td>0.032</td>
<td>0.011</td>
</tr>
</tbody>
</table>

4. Simulation studies

In this section, we conduct simulation studies to investigate the finite sample performance of the proposed method. Let $\mathbf{J}_\text{1}$ be a $4 \times 4$ matrix where every element is equal to 1. The following prior distributions are considered: (i) $\mathbf{\alpha} = 0.5 \times \mathbf{J}_\text{4}$ corresponds to Jeffreys non-informative prior. With complete data, this produces a posterior distribution, where the posterior mean minimizes the Bayes risk under the quadratic loss. (ii) $\mathbf{\alpha} = \mathbf{J}_\text{4}$ gives a posterior mode which is the usual MLE with no missing data. (iii) $\mathbf{\alpha} = 1.5 \times \mathbf{J}_\text{4}$ gives a posterior mode which is the usual MLE after adding 0.5 to each cell count in the $2 \times 2$ table with no missing data.

To eliminate nuisance parameters, the $p$-value we computed is based on the posterior distribution of nuisance parameters. This idea showed up in Guttmann (1967) and Rubin (1984). There is an apparent “double use” of data, first to convert prior into a proper posterior distribution, and then to compute the tail area based on observed value of the test statistic. To circumvent this “double usage” of data, Bayarri and Berger (2000) proposed conditional predictive $p$ value and the partial posterior predictive $p$ value, which follow $\mathcal{U}(0, 1)$ under certain assumptions. It is too complicated for us to use the conditional predictive $p$ value and the partial posterior predictive $p$ value in our set up. To evaluate how close our $p$ value is with the $\mathcal{U}(0, 1)$, in Fig. 1 and Fig. 2, we plot their cumulative distribution functions for four cases (a,b,c,d) and the prior $\mathbf{\alpha} = 0.51_{16}$ used in the simulation studies empirically and put them in Appendix A. It turns out that the cdf of $p_{v1}$ is fairly close to $\mathcal{U}(0, 1)$ and $p_{v2}$ can be anti-conservative most of the time.

To compute the type I error rate, we set up the cell probabilities to satisfy $\theta_2 = \theta_3 - 0.01$ (i.e., $\sum_{i=1}^{4}\pi_{12} = \sum_{i=1}^{4}\pi_{13} - 0.01$) and $\sum_{i=1}^{4}\sum_{j=1}^{4}\pi_{ij} = 1$ as in Table B.1 in the Appendix. We consider three scenarios: no missing data (case a), MAR (case b) and MNAR (case c and d). We consider sample size $N = 30, 50, 100$.

Given $N$ and cell probabilities $\pi$, we generate

\[
(\pi_1, \pi_2, \pi_3, \pi_4, \pi_{12}, \pi_{13}, \pi_{14}, \pi_{23}, \pi_{24}, \pi_{34}, \pi_{123}, \pi_{124}, \pi_{134}, \pi_{234}, \pi_{1234}) \sim \text{Multinomial}(N; \pi_{11}, \ldots, \pi_{14}, \pi_{21} + \pi_{22}, \pi_{23} + \pi_{24}, \pi_{31} + \pi_{32}, \pi_{33} + \pi_{34}, \pi_{41} + \cdots + \pi_{44})
\]

and obtain the observed data $D_{\text{obs}} = \{n_1, \ldots, n_4, m_{12}, m_{34}, m_{13}, m_{24}, m_{1234}\}$. Based on $D_{\text{obs}}$ and $\mathbf{\alpha}$, we generate $G = 5000$ i.i.d. posterior samples $\{\hat{\mathbf{\xi}}_g\}_{g=1}^{G}$, and calculate two Bayesian $p$-values $p_{v1}$ and $p_{v2}$. Repeat this process 10,000 times, we get the empirical type I error rate by counting the percentage of $p$-values that are smaller than 0.05. Bayarri and Berger (2000) showed that $p$-values obtained from the posterior distribution after plugging in corresponding estimates are conservative. The results are summarized in Table 4. The nominal type I error control is 0.05.

From Table 4, we observe that the method based on $p_{v1}$ controls type I error rate conservatively over all scenarios. Method based on $p_{v2}$ has type I error rate close to the nominal 0.05 when there are missing data and has inflated type I error rate when there is no missing data.

We next investigate power. We set up the cell probabilities to satisfy $\theta_2 > \theta_3 - 0.01$ (i.e., $\sum_{i=1}^{4}\pi_{12} > \sum_{i=1}^{4}\pi_{13} - 0.01$, and $\sum_{i=1}^{4}\sum_{j=1}^{4}\pi_{ij} = 1$) as in Table B.2 in the Appendices A and B. We consider three scenarios: no missing data (case a1), MAR (case b1) and MNAR (case c1 and d1). Sample size $N = 30, 50, 100$. The rest are exactly the same as in the calculation of type I error rate. The results are summarized in Table 5.

From Table 5, we observe that the two approaches have very similar power across all priors, sample sizes and missingness mechanism. Jeffreys non-informative prior $\mathbf{\alpha} = 0.5 \times \mathbf{J}_\text{4}$ produces higher power than other priors for both approaches. The proposed methods have higher power under MNAR than under MAR. Power improves with increased sample size.

To investigate the performance of the proposed Bayesian credible intervals for the risk difference, we compute their empirical coverage probabilities (ECPs) and empirical confidence widths (ECWs) through simulations. Table B.3 in the Appendices A and B lists the configuration of cell probabilities $\pi$, which satisfies $\delta = -0.25, 0.25$ and $\sum_{i=1}^{4}\sum_{j=1}^{4}\pi_{ij} = 1$. 

H. Li et al. / Computational Statistics and Data Analysis 127 (2018) 69–81
intervals all contain 0, \( p \) weak evidence to reject the null. On the other hand, under MNAR, \( \alpha = 1 \cdot z_{16} \) intervals are reported in Table 9. We can see that under MAR, when prior \( \alpha \) of \( \pi \) based on posterior credible interval and HPD are similar. Prior \( \alpha \) credible interval and HPD have similar performance across different priors. Empirical coverage probabilities are close to 0.5. A study of medical malpractice width gets smaller. Empirical confidence widths for complete data is the shortest. As sample size increases, the empirical confidence width gets smaller.

5. Two real examples

5.1. A study of medical malpractice

We revisit the study of medical malpractice data example introduced in Section 1. From Table 1, we have \( n_1 = 26, n_2 = 1, n_3 = 5, n_4 = 18, m_{12} = 2, m_{34} = 9, m_{13} = 4, m_{24} = 4, m_{1234} = 0 \). We use the proposed Bayesian \( p \)-values to test hypothesis (2.1). We also analyze this dataset under the MAR assumption. The \( p \)-values are reported in Table 8. Bayesian posterior intervals are reported in Table 9. We can see that under MAR, when prior \( \alpha = 0.5 \times J_4 \), the \( p \)-values are smaller than 0.05 and the Bayesian posterior intervals do not contain 0, suggesting strong evidence to reject the null. Still under MAR, if prior \( \alpha = 1.0 \times J_4 \), the \( p \)-values are borderline significant and Bayesian posterior intervals contain 0. These suggest weak evidence to reject the null. On the other hand, under MNAR, \( p \)-values are greater than 0.05 and the Bayesian posterior intervals all contain 0, consistently showing the we do not reject the null.
Table 6
Empirical coverage probabilities based on 10,000 replications and 5000 sampling samples.

<table>
<thead>
<tr>
<th>N</th>
<th>Prior</th>
<th>( \alpha = 0.5 \cdot 1_{16} )</th>
<th>( \alpha = 1.0 \cdot 1_{16} )</th>
<th>( \alpha = 1.5 \cdot 1_{16} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPD</td>
<td>Credible interval</td>
<td>HPD</td>
<td>Credible interval</td>
</tr>
<tr>
<td>30</td>
<td>Case ( a_2 )</td>
<td>0.953 (0.962)</td>
<td>0.980 (0.982)</td>
<td>0.987 (0.987)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.992 (0.992)</td>
<td>0.996 (0.995)</td>
<td>0.956 (0.956)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>0.998 (0.998)</td>
<td>0.990 (0.991)</td>
<td>0.951 (0.952)</td>
</tr>
<tr>
<td>50</td>
<td>Case ( a_2 )</td>
<td>0.947 (0.954)</td>
<td>0.972 (0.974)</td>
<td>0.981 (0.982)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.998 (0.997)</td>
<td>0.992 (0.992)</td>
<td>0.978 (0.977)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>1.000 (1.000)</td>
<td>0.993 (0.993)</td>
<td>0.982 (0.981)</td>
</tr>
<tr>
<td>100</td>
<td>Case ( a_2 )</td>
<td>0.946 (0.951)</td>
<td>0.950 (0.953)</td>
<td>0.967 (0.965)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.999 (0.999)</td>
<td>1.000 (1.000)</td>
<td>0.997 (0.996)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>1.000 (1.000)</td>
<td>0.998 (0.998)</td>
<td>0.996 (0.996)</td>
</tr>
<tr>
<td>400</td>
<td>Case ( a_2 )</td>
<td>0.952 (0.954)</td>
<td>0.952 (0.951)</td>
<td>0.962 (0.963)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.999 (0.999)</td>
<td>1.000 (1.000)</td>
<td>0.997 (0.996)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>1.000 (1.000)</td>
<td>1.000 (1.000)</td>
<td>1.000 (1.000)</td>
</tr>
</tbody>
</table>

Table 7
Empirical confidence widths based on 10,000 replications and 5000 sampling samples.

<table>
<thead>
<tr>
<th>N</th>
<th>Prior</th>
<th>( \alpha = 0.5 \cdot 1_{16} )</th>
<th>( \alpha = 1.0 \cdot 1_{16} )</th>
<th>( \alpha = 1.5 \cdot 1_{16} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPD</td>
<td>Credible interval</td>
<td>HPD</td>
<td>Credible interval</td>
</tr>
<tr>
<td>30</td>
<td>Case ( a_2 )</td>
<td>0.437 (0.438)</td>
<td>0.400 (0.401)</td>
<td>0.371 (0.373)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.510 (0.512)</td>
<td>0.445 (0.446)</td>
<td>0.402 (0.403)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>0.513 (0.515)</td>
<td>0.446 (0.448)</td>
<td>0.402 (0.404)</td>
</tr>
<tr>
<td>50</td>
<td>Case ( a_2 )</td>
<td>0.357 (0.359)</td>
<td>0.334 (0.336)</td>
<td>0.318 (0.319)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.449 (0.451)</td>
<td>0.393 (0.395)</td>
<td>0.358 (0.360)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>0.449 (0.450)</td>
<td>0.392 (0.394)</td>
<td>0.358 (0.359)</td>
</tr>
<tr>
<td>100</td>
<td>Case ( a_2 )</td>
<td>0.263 (0.264)</td>
<td>0.254 (0.255)</td>
<td>0.246 (0.247)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.386 (0.388)</td>
<td>0.333 (0.335)</td>
<td>0.304 (0.305)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>0.386 (0.387)</td>
<td>0.334 (0.335)</td>
<td>0.305 (0.306)</td>
</tr>
<tr>
<td>400</td>
<td>Case ( a_2 )</td>
<td>0.137 (0.137)</td>
<td>0.135 (0.136)</td>
<td>0.134 (0.134)</td>
</tr>
<tr>
<td></td>
<td>Case ( b_2 )</td>
<td>0.385 (0.388)</td>
<td>0.333 (0.335)</td>
<td>0.304 (0.305)</td>
</tr>
<tr>
<td></td>
<td>Case ( c_2 )</td>
<td>0.323 (0.324)</td>
<td>0.270 (0.268)</td>
<td>0.236 (0.237)</td>
</tr>
</tbody>
</table>

Table 8
Results for testing (2.1) in the study of medical malpractice example.

<table>
<thead>
<tr>
<th>Prior</th>
<th>( \alpha = 0.5 \cdot 1_{16} )</th>
<th>( \alpha = 1.0 \cdot 1_{16} )</th>
<th>( \alpha = 1.5 \cdot 1_{16} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_{11} )</td>
<td>( p_{12} )</td>
<td>( p_{11} )</td>
</tr>
<tr>
<td>p-value under NMAR</td>
<td>0.899</td>
<td>0.109</td>
<td>0.899</td>
</tr>
<tr>
<td>p-value under MAR</td>
<td>0.024</td>
<td>0.038</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Table 9
95% Bayesian intervals for \( \delta \) in the study of medical malpractice example.

<table>
<thead>
<tr>
<th>Missing mechanism</th>
<th>Prior</th>
<th>Type of Bayesian intervals</th>
<th>( \delta ) Value</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPD</td>
<td>Credible interval</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Width</td>
<td>Lower</td>
</tr>
<tr>
<td>Missing mechanism</td>
<td>Prior</td>
<td>HPD</td>
<td>Credible interval</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Width</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 0.5 \cdot 1_{16} )</td>
<td>-0.249</td>
<td>0.0507</td>
<td>0.299</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 1.0 \cdot 1_{16} )</td>
<td>-0.228</td>
<td>0.045</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 1.5 \cdot 1_{16} )</td>
<td>-0.212</td>
<td>0.046</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 0.5 \cdot 1_{16} )</td>
<td>-0.218</td>
<td>-0.0320</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 1.0 \cdot 1_{16} )</td>
<td>-0.172</td>
<td>0.0133</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 1.5 \cdot 1_{16} )</td>
<td>-0.160</td>
<td>0.014</td>
<td>0.174</td>
</tr>
</tbody>
</table>

5.2. A crime survey example

We illustrate the proposed method with data from a crime survey study, which was reported in Kadane (1985) and Tian et al. (2003). We have \( n_1 = 392, n_2 = 55, n_3 = 76, n_4 = 38, m_{12} = 33, m_{34} = 9, m_{13} = 31, m_{24} = 7, m_{1234} = 115 \). The \( p \)-values are summarized in Table 10 and Bayesian posterior intervals are summarized in Table 11. Under MAR, \( p \)-values show evidence to reject the null yet the Bayesian posterior intervals all contain 0 regardless of the prior. Under NMAR,
6. Concluding remarks

In this paper, we study non-inferiority testing with paired data under non-ignorable missing mechanism. We present two Bayesian p-value approaches and provide Bayesian credible intervals and highest posterior density intervals for the risk difference. Numerical studies show that the proposed methods can control type I error rate and have decent power. Empirical coverage probabilities always exceed the nominal 95%. We find Jeffreys non-informative prior consistently has good performance and recommend its usage in practice.

Acknowledgments

The authors would like to thank the editor, the associate editor and two referees for insightful comments that help to streamline the paper. The research of Hui-qiong Li is partially supported by the Natural Science Foundation of China (11561075) and grants from the Science Foundation of Yunnan Province (2016FB005). Guo-liang Tian is partly supported by the Natural Science Foundation of China (11771199).

Appendix A. Cumulative distribution function of $p_{v1}$ and $p_{v2}$

See Figs. 1 and 2.

Appendix B. This appendix contains cell probabilities to generate data in the simulation studies

See Tables B.1–B.3.
Fig. 1. cdf under the prior $\alpha = 0.51_{16}$.

Table B.2
Cell probabilities for calculating the power.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$</td>
</tr>
<tr>
<td>b1</td>
<td>$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0)$</td>
</tr>
<tr>
<td>c1</td>
<td>$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0)$</td>
</tr>
<tr>
<td>d1</td>
<td>$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0)$</td>
</tr>
</tbody>
</table>
Fig. 2. cdf under the prior $\alpha = 0.5_{16}$.

Table B.3
Cell probabilities for empirical coverage probabilities and empirical confidence widths.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\delta$</th>
<th>$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a2</td>
<td>$0$</td>
<td>$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\right)^\top$</td>
</tr>
<tr>
<td>b2</td>
<td>$-0.25$</td>
<td>$\left(\frac{1}{2}, 0, \frac{1}{2}, \frac{1}{4}, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0\right)^\top$</td>
</tr>
<tr>
<td>c2</td>
<td>$0.25$</td>
<td>$\left(\frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)^\top$</td>
</tr>
</tbody>
</table>
References