Published in final edited form as:

Stat Methods Med Res. ; : 962280217702538. doi:10.1177/0962280217702538.

Exact Bayesian p-values for a Test of Independence in a 2×2 Contingency Table with Missing Data

Yan Lin¹, Stuart Lipsitz², Debajyoti Sinha³, Garrett Fitzmaurice⁴, and Steven Lipshultz⁵

¹M.D. Anderson Cancer Center, Houston, Texas USA

²Brigham and Women's Hospital, Boston, MA

³The Florida State University, Tallahassee, Florida, USA

⁴McLean Hospital, Belmont, MA, USA

⁵Wayne State University Medical School, Detroit, MI, USA

Abstract

Altham (Altham PME. Exact Bayesian analysis of a 2×2 contingency table, and Fisher's "exact" significance test. *J. R. Statist. Soc. B* 1969; 31; 261–269) showed that a one-sided p-value from Fisher's exact test of independence in a 2×2 contingency table is equal to the posterior probability of negative association in the 2×2 contingency table under a Bayesian analysis using an improper prior. We derive an extension of Fisher's exact test p-value in the presence of missing data, assuming the missing data mechanism is ignorable (i.e., missing at random or completely at random). Further, we propose Bayesian p-values for a test of independence in a 2×2 contingency table with missing data using alternative priors; we also present results from a simulation study exploring the Type I error rate and power of the proposed exact test p-values. An example, using data on the association between blood pressure and a cardiac enzyme, is presented to illustrate the methods.

Keywords

Dirichlet prior; Fisher's exact test; missing at random; missing completely at random; quasi-Monte Carlo integration

1. Introduction

A frequently used statistical test in biomedical applications is the test of independence between row and column variables in a 2×2 contingency table. Even in studies where more complicated analyses are subsequently performed, initial summaries in terms of bivariate analyses are regularly reported. With small cell counts in a 2×2 contingency table, Fisher's exact test is often preferred. However, a frequently occurring complication in many studies

is missing data, e.g., either the row or column variable is missing for some of the experimental units.

An example of a 2×2 contingency table with missing data arose in a study of 269 ill children examined in the emergency room at Boston's Children's Hospital¹. These children are at risk, but rarely screened, for myocardial injury; Lipshultz et al.¹ addressed this issue by measuring cardiac enzymes in these children. One of the cardiac enzymes measured was myoglobin; values of myoglobin > 85 ng/ml are an indication of myocardial injury. The study investigators were interested in whether simple and less costly measures of heart function, such as diastolic blood pressure, are associated with levels of myoglobin. In particular, the investigators wanted to test if abnormally low diastolic blood pressure (DBP), defined as $\text{DBP} < 50\text{mmHg}$, is associated with abnormal levels of myoglobin. Data from this study are presented in the 2×2 contingency table in Table 1. We note here that although the underlying variables in the 2×2 table are continuous, the study cardiologist was interested in the association between the dichotomized versions of the variables that represent abnormal levels. As seen in Table 1, 50 children had DBP measured, but did not have myoglobin measured; and 145 children had myoglobin measured, but did not have DBP measured. Further, these 195 patients with missing data constituted 72.5% of the patients in the study. In discussing the reasons for missingness with the study investigators, myoglobin and/or DBP was less likely to be measured on patients who appear to be less ill, so the data are unlikely to be missing completely at random (MCAR)². With such a high percentage of missing data thought not to be MCAR, a so-called "complete case analysis" using only the 74 patients with measures of both variables could give biased results. Even if there were no missing data, a large sample test for independence, such as Pearson's chi-squared test, could give spurious results since the frequency of abnormal myoglobin is so low; instead an exact test, such as Fisher's exact test, would be more appropriate for these data.

With missing data that are assumed to be missing at random (MAR)², likelihood ratio or score statistics³ can be used to test for independence in incomplete 2×2 contingency tables. However, just as is the case with complete data, these likelihood-based methods rely on asymptotic properties and can have high Type I error rates when the contingency table is sparse. An extension of Fisher's exact test that accounts for missing data assumed to be MAR would be more appropriate in these settings; to date, such a test has not been developed.

Here, we propose an extension of Fisher's exact test to handle missing data by extending the Bayesian justification used by Altham⁴ to obtain Fisher's exact p-values. In particular, in an exact Bayesian analysis using a specific Dirichlet prior distribution for the cell probabilities of the 2×2 table, Altham⁴ showed that the "Bayesian p-value", the posterior probability of negative association in the contingency table, is identical to the one-sided Fisher's exact test p-value for the alternative hypothesis of positive association. In general, the Bayesian p-value is a tail probability based on a posterior distribution, and was also discussed by Altham⁵ and by Cassella and Berger⁶. Similarly, our proposed extension of Fisher's exact test is based on the posterior probabilities using the same prior as in Altham⁴, but with the likelihood based on data from subjects with both variables fully observed as well as data from subjects with only one of the two variables observed. Further, we also explore three

alternative Bayesian p-values using different Dirichlet prior distributions for the cell probabilities of the 2×2 contingency table. Use of these other priors allows us to explore how sensitive the Type I error and power are to different choices of Dirichlet priors. The proposed methods are an extension of the Bayesian exact p-value approach for McNemar's test of correlated proportions with missing data discussed in Lin et al.⁷

2. Derivation of exact p-value with missing data

Let $\theta_{ij}(i,j=1,2)$ be the cell probabilities of the 2×2 contingency table, where i indexes rows, and j indexes columns. With missing data, the observed data can be summarized in the form of a 2×2 table with supplemental margins, as shown in Table 2. In this table, Y_{ij} denotes the number of subjects who are fully observed on both the row and column variables, with response level i on the row variable and level j on the column variable. Also, Z_{i+} denotes the number of subjects with response level i on the row variable who are missing the column variable; and U_{+j} denotes the number of subjects with response level j on the column variable who are missing the row variable. We denote the observed data by $D = [Y_{11}, Y_{12}, Y_{21}, Y_{22}, Z_{1+}, Z_{2+}, U_{+1}, U_{+2}]$.

With no missing data, i.e., when $Z_{i+} = U_{+j} = 0$ for all (i,j) , the observed data, denoted by D_c , consists only of $[Y_{11}, Y_{12}, Y_{21}, Y_{22}]$, and the Y_{ij} 's follow a multinomial distribution with cell probabilities θ_{ij} 's. With data missing at random, the likelihood function³ based on all of the observed data, D , in Table 2 is proportional to

$$L(\theta | D) \propto \theta_{11}^{y_{11}} \theta_{12}^{y_{12}} \theta_{21}^{y_{21}} \theta_{22}^{y_{22}} \theta_{1+}^{z_1+} \theta_{2+}^{z_2+} \theta_{+1}^{u_{+1}} \theta_{+2}^{u_{+2}}, \quad (1)$$

where the "+" subscript in the θ_{ij} 's represents summation over the corresponding index⁸ of the multinomial probabilities, θ_{ij} 's, in the likelihood; Z_{i+} and U_{+j} contribute binomial distributions with probabilities θ_{i+} and θ_{+j} , respectively.

First, we discuss the Bayesian p-value when there are no missing data, i.e., where the observed data are $D_c = [Y_{11}, Y_{12}, Y_{21}, Y_{22}]$. Altham⁴ derived a relationship between the posterior probability of negative association in a 2×2 contingency table and the p-value of Fisher's exact test for the one-sided alternative hypothesis of positive association, $H_A: OR > 1$, where $OR = \theta_{11} \theta_{22} / (\theta_{12} \theta_{21})$ is the odds ratio in the 2×2 contingency table. In particular, suppose the joint prior density of $\theta = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ is Dirichlet, with parameters $\mathbf{a} = (a_{11}, a_{12}, a_{21}, a_{22})$, given by the density kernel

$$\pi(\theta) \propto \theta_{11}^{a_{11}-1} \theta_{12}^{a_{12}-1} \theta_{21}^{a_{21}-1} \theta_{22}^{a_{22}-1}, \quad (2)$$

and denoted by $\theta \sim \text{Dir}(\mathbf{a})$. Then, if there are no missing data, the joint posterior distribution of θ is proportional to

$$p(\theta | D_c) \propto \theta_{11}^{y_{11} + \alpha_{11} - 1} \theta_{12}^{y_{12} + \alpha_{12} - 1} \theta_{21}^{y_{21} + \alpha_{21} - 1} \theta_{22}^{y_{22} + \alpha_{22} - 1},$$

and this posterior is proper only when all $y_{ij} + \alpha_{ij} > 0$ for all (i,j) . For the improper Dirichlet prior with $\alpha_{12} = \alpha_{21} = 1$ and $\alpha_{11} = \alpha_{22} = 0$ in (2), the posterior probability of negative association given the observed data D_c is

$$\Pr[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1 | D_c] = \iiint C(y + \alpha) \theta_{11}^{y_{11} + \alpha_{11} - 1} \theta_{12}^{y_{12} + \alpha_{12} - 1} \theta_{21}^{y_{21} + \alpha_{21} - 1} \theta_{22}^{y_{22} + \alpha_{22} - 1} I_{[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1]} d\theta_{11} d\theta_{12} d\theta_{21},$$

where $C(\eta)$ is the normalizing constant $\Gamma\left(\sum_{i,j} \eta_{ij}\right) \prod_{i,j} \Gamma(\eta_{ij})$, $\eta_{ij} = y_{ij} + \alpha_{ij}$, $\Gamma(\cdot)$ is a gamma function, $I[\cdot]$ is an indicator function, $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$, and the integral is over all possible values of $(\theta_{11}, \theta_{12}, \theta_{21})$ such that $0 < \theta_{jk} < 1$ and $0 < \theta_{11} + \theta_{12} + \theta_{21} < 1$. Altham⁴ showed that this posterior probability of negative association, given the observed data D_c and the improper Dir(0,1,1,0) prior for θ , can be written as a finite sum

$$\Pr[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1 | D_c, \alpha = (0, 1, 1, 0)] = \sum_{z = \max(\eta_{12} - \eta_{21}, 0)}^{\eta_{12} - 1} \binom{\eta_{+2} - 1}{s} \binom{\eta_{+1} - 1}{\eta_{+1} - 1 - s} \binom{\eta_{+} - 2}{\eta_{2+} - 1} \quad (3)$$

provided that $y_{11} > 0$, $y_{22} > 0$, where again $\eta_{ij} = y_{ij} + \alpha_{ij}$. For the case when $y_{11} = 0$ or $y_{22} = 0$, the resulting posterior is improper with $\Pr[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1 | D_c, \alpha = (0, 1, 1, 0)] \rightarrow \infty$, and thus the posterior probability that $OR < 1$ cannot be computed for the Dir(0,1,1,0). We note here that the right side of (3) identically equals Fisher's exact test p-value for the one-sided alternative $H_A: OR > 1$, and even though the left hand side of (3) approaches infinity when at least one of y_{11} , y_{22} is 0, the right hand side of (3) (the one-sided Fisher's exact p-value) exists and equals 1 when y_{11} and/or y_{22} is 0.

As described by Altham⁴, if one is interested in the alternative that $H_A: OR < 1$, then we can use an improper Dir(1,0,0,1) prior for θ to calculate the one-sided Bayesian p-value $\Pr[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) > 1 | D_c, \alpha = (1, 0, 0, 1)]$. If both $y_{12} > 0$ and $y_{21} > 0$, the resulting posterior probability that the $OR > 1$ identically equals Fisher's exact test p-value for the one-sided alternative $H_A: OR < 1$. If either $y_{12} = 0$ or $y_{21} = 0$, the resulting posterior is improper with $\Pr[\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) > 1 | D_c, \alpha = (1, 0, 0, 1)] \rightarrow \infty$, and thus the posterior probability that $OR > 1$ cannot be computed for the Dir(1,0,0,1). Again, we note that even though the posterior probability approaches infinity when at least one of y_{12} , y_{21} is 0, the one-sided Fisher's exact p-value exists and equals 1 when y_{12} and/or y_{21} is 0.

To overcome the problems with these $\text{Dir}(0,1,1,0)$ and $\text{Dir}(1,0,0,1)$ priors with 0 counts on the diagonal or off diagonal, we formulate alternative Bayesian p-values by calculating the posterior probability that $\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1$ and the posterior probability that $\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) > 1$ using other specifications $[a_{11}, a_{10}, a_{01}, a_{00}]$ in which all a_{ij} 's are positive. In this case, all posterior distributions will be proper since $\eta_{ij} = y_{ij} + a_{ij} > 0$ for all i, j . In particular, we consider the priors $a_{ij} = a = 0.5$ for all i and j ; $a_{ij} = a = 1.0$ for all i and j ; and $a_{ij} = a = 1.5$ for all i and j .

For intuition, with no missing data, and when there is interest in estimating the odds ratio of the 2×2 contingency table, $a=0.5$ corresponds to Jeffreys non-informative prior, which gives a mean of the posterior distribution of the cell probabilities which has minimum expected square error over the sampling distribution of the data (the frequentist mean square error of the Bayesian estimate)⁹; $a=1.0$ leads to a mode of the posterior distribution of the cell probabilities which is the usual MLE; $a=1.5$ leads to a mode of the posterior distribution of the cell probabilities which is the usual MLE after adding 0.5 to each cell count.

We also note here that Altham⁴ derived the one-sided Fisher's exact test p-values by fixing the row totals so that we have two independent binomial samples with appropriate beta priors. Further, Altham⁴ did not consider a Bayesian approach to obtain a 2-sided alternative. As discussed in Fleiss et al.¹⁰, there are two possibilities for 2-sided p-values for Fisher's exact test, both of which have the correct Type I error. One possible two-sided p-value is calculated by summing all hypergeometric probabilities, conditional on the observed row and column totals, that are less than or equal to the observed table probability. The other two-sided p-value is simply calculated as two times the minimum of the two one-sided p-values:

$$p = 2 \min \{ \Pr [\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1 \mid D_c, \alpha = (0, 1, 1, 0)], \Pr [\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) > 1 \mid D_c, \alpha = (1, 0, 0, 1)] \}.$$

(4)

The two-sided p-value in (4) is the one which we extend to the case with ignorable (MAR or MCAR) missing data. For all priors discussed here, except possibly the priors $\text{Dir}(0,1,1,0)$ and $\text{Dir}(1,0,0,1)$, both one sided-probabilities in (4) will be in $(0,1]$. For the priors $\text{Dir}(0,1,1,0)$ and $\text{Dir}(1,0,0,1)$, the two sided p-value in (4) always gives a P-value in $(0,1)$ even when there are 0 counts. This is because, with 0 counts, one of the one-sided posterior probabilities does not exist (not bounded), and the other posterior probability will be in $(0,1]$; thus, 2 times the minimum of the two one-sided P-values will be in $(0,1)$.

Next, we discuss the Bayesian p-value with missing data. Using the same $\text{Dir}(\alpha)$ prior given in (2) and assuming an ignorable missing data mechanism for the data D in Table 2, the likelihood is given by (1), and the joint posterior distribution of θ is proportional to:

$$p(\theta | D) \propto \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \theta_{1+}^{z_1} \theta_{2+}^{z_2} \theta_{+1}^{u+1} \theta_{+2}^{u+2}, \quad (5)$$

which is proper if and only if each $y_{ij} + a_{ij} > 0$. In particular, this “all data” posterior in (5) is bounded above by the complete case posterior,

$$\begin{aligned} & \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \theta_{1+}^{z_1} \theta_{2+}^{z_2} \theta_{+1}^{u+1} \theta_{+2}^{u+2} \\ & \leq \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1}, \end{aligned} \quad (6)$$

and it is bounded below by two functions that are similar to the complete case posterior,

$$\begin{aligned} & \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \theta_{1+}^{z_1} \theta_{2+}^{z_2} \theta_{+1}^{u+1} \theta_{+2}^{u+2} \\ & \geq \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \end{aligned} \quad (7)$$

and

$$\begin{aligned} & \theta_{11}^{y_{11}+a_{11}-1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \theta_{1+}^{z_1} \theta_{2+}^{z_2} \theta_{+1}^{u+1} \theta_{+2}^{u+2} \\ & \geq \theta_{11}^{y_{11}+z_1+u+1} \theta_{12}^{y_{12}+a_{12}-1} \theta_{21}^{y_{21}+a_{21}-1} \theta_{22}^{y_{22}+a_{22}-1} \theta_{+2}^{z_2+u+2}. \end{aligned} \quad (8)$$

The “all data” posterior with missing data (left hand side of all above three inequalities (6), (7), and (8)) will be proper if and only if the right hand side of each of the above three inequalities (6), (7), and (8) is integrable. This is also equivalent to the set of the conditions that $y_{ij} + a_{ij} > 0$ for all i and j . For all priors discussed here other than the Fisher’s exact priors, $a_{ij} > 0$ for all i and j , and thus the “all data” posterior with missing data will be proper. For Fisher’s exact priors, we must separately consider the Dir(0,1,1,0) and Dir(1,0,0,1) priors. For Dir(0,1,1,0), the right hand side of (7) is not integrable if either y_{11} or y_{22} is 0 (i.e., either of these complete case cell counts are 0), and thus the left hand side of (7), the “all data” posterior, is also not integrable; similarly, for Dir(1,0,0,1), the right hand side of (8) is not integrable if either y_{12} or y_{21} is 0, and thus the left hand side of (8), the “all data” posterior, is also not integrable. Further, for Dir(0,1,1,0), when both y_{11} and y_{22} are positive, the right hand side of (6) and right hand side of (7) are integrable, so that the “all data” posterior is also integrable; similarly, for Dir(1,0,0,1) when both y_{12} and y_{21} are positive for Dir(1,0,0,1), the right hand side of (6) and right hand side of (8) are integrable, so that the “all data” posterior is also integrable.

Thus, a necessary and sufficient condition for the “all data” posterior using the Fisher Exact priors to be proper is that both y_{11} and y_{22} are positive for Dir(0,1,1,0) and both y_{12} and y_{21}

are positive for $\text{Dir}(1,0,0,1)$. In particular, both the “all data” and “Complete case” posteriors using the Fisher Exact priors will be proper if and only if both y_{11} and y_{22} are positive for $\text{Dir}(0,1,1,0)$ and both y_{12} and y_{21} are positive for $\text{Dir}(1,0,0,1)$.

Our proposed two-sided p-value based on (5) is

$$p = 2 \min \{ \Pr [\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) < 1 \mid D, \alpha = (0, 1, 1, 0)], \Pr [\theta_{11}\theta_{22}/(\theta_{12}\theta_{21}) > 1 \mid D, \alpha = (1, 0, 0, 1)] \} .$$

(9)

As with complete cases, for all priors discussed here, except possibly the priors $\text{Dir}(0,1,1,0)$ and $\text{Dir}(1,0,0,1)$, both one sided-probabilities in (4) will be in $(0,1]$. Even though both one-sided posteriors will not be proper for the priors $\text{Dir}(0,1,1,0)$ and $\text{Dir}(1,0,0,1)$ with a 0 cell in the complete case table, one of the two one-sided posteriors will be proper, and thus the minimum of the two probabilities in (9) will be in $(0,1)$.

Unfortunately, in the presence of missing data, the one-sided probabilities in (9) do not have a closed form expression as in (3), and require a numerical integration technique. Since the dimension of the integration is low (3 dimensions), the posterior probabilities (Bayesian p-values) can be computed using direct numerical integration; we used quasi-Monte Carlo integration^{11, 12} with 100,000 quadrature points for the motivating data example and simulation study. The method is implemented in a SAS macro that is available upon request from the first author. With data in contingency table form as in Table 2, the SAS macro takes approximately 0.2 seconds of CPU time to calculate a p-value on a less than state-of-the-art, 32-bit Windows 7 PC with an Intel third Generation Core i5-3570 processor and 4GB RAM. Alternatively, Markov Chain Monte Carlo (MCMC), via the software WINBUGS, can be used.

3. Motivating Example

In this section, we illustrate the key results of this paper using data from the study¹ of diastolic blood pressure (DBP) and myoglobin discussed in the Introduction and given in Table 1. Recall that there are 195 children with missing data, which constituted 72.5% of the children in the study. Here, we assume the missing data on these 195 children are missing at random² based on the reasons for missingness suggested by the investigators of the original study¹. Our interest here centers on whether there is an association between abnormal levels of myoglobin and DBP. Because the frequency of abnormal myoglobin is so low, an exact test of independence is appropriate.

Before considering an exact test for association, it is of interest to explore the missing data mechanism that might be generating the missing data. A somewhat informal way¹³ to assess if the data are missing completely at random (so that an analysis restricted to complete cases only will give appropriate results) is to compare the marginal proportions in the 2×2 table

of complete cases to the marginal proportions in those missing a row or column variable. In the complete cases, myoglobin is low $2/72=2.78\%$ of the time, whereas in subjects with only myoglobin observed, myoglobin is low $1/145=0.69\%$ of the time. Fisher's exact test comparing these two proportions is not significant ($p=0.256$), which suggests that the missing data on myoglobin may be missing completely at random. In the complete cases, DBP is low $25/74=33.8\%$ of the time, whereas in subjects with only DBP observed, DBP is low $1/50=2.0\%$ of the time. Fisher's exact test comparing these two proportions is significant ($p<.0001$), suggesting that the missing DBP data are not missing completely at random.

Here, our main interest is in the two-sided alternative that $H_A: OR \neq 1$, i.e., that abnormal levels of myoglobin and DBP are associated. For completeness, in Table 3, we present the one and two-sided p-values for the different priors, using all of the available data as well as data only on the complete cases. From Table 3 we see that the p-value for the two-sided alternative with the 'Fisher's exact prior' using all of the data is borderline significant ($p<0.10$), whereas the complete cases p-value is much further from the 10% level. For all other priors, the p-value using all of the data is significant at a 5% level, whereas none of the complete case p-values are significant at the 5%-level. We also see that among all priors using both complete cases and all of the data, the 'Fisher's exact prior' gives the most conservative p-value.

This example nicely illustrates how different approaches for handling missingness can lead to different, and possibly conflicting results in the analysis of contingency table data. Further, although by a very slight margin $\alpha=1.5$ produces the smallest p-value (based on all of the data) in this example; we caution that this results may not generalize. To further explore properties of the proposed method, we performed a simulation study that compared the different approaches; the results of the simulation study are described in the following section.

4. Simulation Study

To explore the properties of the proposed p-values, we performed a simulation study for a 2×2 contingency table with missing data similar to that found in Table 2. We chose a total sample size of 60 and fixed the number of simulation replications at 2000 for a given configuration of the data. We considered the scenario in which the cell probabilities $\theta_{11} = 0.25$ and $\theta_{22} = 0.25$; θ_{11} varied from 0.25 to 0.45 in increments of 0.05, and correspondingly, θ_{12} varied from 0.25 to 0.05, giving odds ratios that range from 1 to 9. In the simulations, we used the ignorable missing at random mechanism discussed in Chen and Fienberg¹⁴.

To describe the missing data model of Chen and Fienberg¹⁴, it is convenient to introduce row and column outcome binary variables and missing data indicator variables. In particular, let V_1 be the 'row' binary random variable, which can take on values 0 or 1, and let V_2 be the 'column' binary random variable, which can also take on values 0 or 1. We define the missing data indicator R_k , where $R_k = 1$ if V_k is missing; and $R_k = 0$ if V_k is observed, for $k=1,2$. The conditional multinomial distribution of the missing data indicators (R_1, R_2) given

(V_1, V_2) is called the ‘missing data mechanism’. Under a missing at random ‘missing data mechanism’, the probability that V_1 is observed and V_2 is missing only depends on the observed value of V_1 ,

$$\Pr(R_1 = 0, R_2 = 1 \mid V_1 = v_1, V_2 = v_2) = v_1 \pi_1 + (1 - v_1) \pi_2,$$

with probabilities (π_1, π_2) . Similarly, the probability that V_2 is observed and V_1 is missing only depends on the observed value of V_2 ,

$$\Pr(R_1 = 1, R_2 = 0 \mid V_1 = v_1, V_2 = v_2) = v_2 \omega_1 + (1 - v_2) \omega_2,$$

with probabilities (ω_1, ω_2) . Since, under the MAR missing data mechanism proposed by Chen and Fienberg¹⁴, the probability that both V_1 and V_2 are missing is 0, the probability that both V_1 and V_2 are observed is $1 - [v_1 \pi_1 + (1 - v_1) \pi_2] - [v_2 \omega_1 + (1 - v_2) \omega_2]$, which depends on the values of both V_1 and V_2 . We note that if there are subjects missing both row and column variables (which does not occur in our dataset), one could add a constant parameter to the last cell in Table 4, and subtract this constant in the square brackets in the complete case 2×2 table.

This MAR missing data mechanism of Chen and Fienberg¹⁴ satisfies the assumption that the probability of a row or column variable being missing depends only on observed data. Under this MAR model, the multinomial probabilities of the joint distribution of (R_1, V_1, R_2, V_2) are shown in Table 4. A given missing data configuration entails specification of the four probabilities $(\pi_1, \pi_2, \omega_1, \omega_2)$. In the simulations, we specified three sets of $(\pi_1, \pi_2, \omega_1, \omega_2)$. The first set is $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .45, .05)$; the second set is $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .05, .45)$; and the third set is $(\pi_1, \pi_2, \omega_1, \omega_2) = (.25, .25, .25, .25)$. Note that the third set is actually ‘missing completely at random’³ (MCAR), since $\Pr(R_1 = r_1, R_2 = r_2 \mid V_1 = v_1, V_2 = v_2) = \Pr(R_1 = r_1, R_2 = r_2)$, independent of any observed data (V_1, V_2) .

We note here that this MAR model/assumption is convenient to make for a 2×2 table; in practice, the researcher will typically have other observed variables in the dataset, and a more plausible MAR assumption might be that the probability that a variable is missing depends on these other observed variables. However, to make inference under such an MAR mechanism, one would need to set-up and estimate a model for the missing data given the observed data (other observed variables), which could be unstable with small cell counts in the 2×2 table of interest.

Before presenting and discussing the results of the simulations, we can give some insight into why the Type I error can be elevated under MAR when restricting the analysis to ‘complete cases’, i.e., subjects with $(R_1 = 0, R_2 = 0)$. From Table 4, the odds ratio for the 2×2 table of complete cases is

$$OR_{CC} = \frac{(1 - \pi_1 - \omega_1)(1 - \pi_2 - \omega_2)\theta_{11}\theta_{22}}{(1 - \pi_1 - \omega_2)(1 - \pi_2 - \omega_1)\theta_{112}\theta_{21}} = \frac{(1 - \pi_1 - \omega_1)(1 - \pi_2 - \omega_2)}{(1 - \pi_1 - \omega_2)(1 - \pi_2 - \omega_1)}OR, \quad (10)$$

where OR is the true odds ratio between V_1 and V_2 . Under the null H_0 : $OR = 1$, for the MAR mechanism $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .45, .05)$, the complete case OR_{CC} in (10) equals 2.78, and for the MAR mechanism $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .05, .45)$, the complete case OR_{CC} in (10) equals 0.36. Thus, for both MAR mechanisms, a complete case test for independence could give high Type 1 error since OR_{CC} is far from 1. Under the null H_0 : $OR = 1$, for the MCAR mechanism $(\pi_1, \pi_2, \omega_1, \omega_2) = (.25, .25, .25, .25)$, the complete case OR_{CC} in (10) equals 1, so a complete case test should give the correct Type 1 error. Since using all of the data and the complete cases only gives correct Type I error under MCAR, the MCAR mechanism will allow us to explore the additional power gained from using all of the available data (versus the complete cases only).

Another intuitive way to think about biases that can result from this MAR missing data mechanism described in Table 4 is related to Simpson's paradox. Using Bayes' rule, one can show that the odds ratio for the subjects with $(R_1 = 0, R_2 = 1)$, i.e., based on the probabilities $\Pr(V_1 = v_1, V_2 = v_2 | R_1 = 0, R_2 = 1)$, equals the odds ratio that is of interest OR ; similarly, the odds ratio for the subjects with $(R_1 = 1, R_2 = 0)$, equals the odds ratio that is of interest OR . Then, when marginalising over the three subpopulations of complete cases and subjects with only row or column variables, two of which have the odds ratio OR (those missing row or column variables) and one of which has the odds ratio $OR_{CC} \neq OR$, we obtain a population odds ratio that equals the OR of interest.

In the simulations, we explored the Type I error rate and power for a 5% significance level test, under the three missing data mechanisms for the two-sided alternative H_A : $OR \neq 1$. We calculated the two-sided p-value using all of the data D and also using only the complete cases D_C ; the results are displayed in Table 5. As discussed above, for the two MAR mechanisms, a complete cases analysis could give high Type 1 error since the OR_{CC} is far from 1 under the null. As can be seen in Table 5, the Type I error is high for all complete case tests, whereas the Type I error is at most 5% using all of the available data. Note that, because of the conservativeness of Fisher's exact test, a 'high' Type 1 error means one of approximately 7%. For the first MAR mechanism in which $OR_{CC} > 1$ under both the null and alternative, the Bayesian p-value based on all of the data with $\alpha = 0.5$ gives the highest power (along with correct Type 1 error rate), with the Fisher's exact prior giving the lowest power. For the second MAR mechanism we see that the rejection percentages (estimated power) for complete cases actually decreases as the true OR ranges from 1 to 2.33 because OR_{CC} gets closer to 1 in this interval (specifically, OR_{CC} moves from 0.36 to 0.84 in this interval). Again, using all of the data, the Type I error is at most 5% across the different priors; the Fisher's exact prior is the most conservative, and the Bayesian prior with $\alpha = 0.5$ gives the highest power. This second MAR mechanism is perhaps most similar to the mechanism generating the data for the motivating data example analyzed in Section 3.

Under the MCAR mechanism in Table 5, we see that for all practical purposes no power is gained by including the patients with missing data in the analysis. Interestingly, this agrees with asymptotic results found in Lipsitz and Fitzmaurice¹⁵, where they show that the complete case score statistic (Pearson's chi-squared) for testing independence in a 2×2 contingency table is fully efficient when the missing data are MCAR.

5. Conclusion

In summary, we have proposed exact Bayesian p-values for testing independence between two dichotomous variables when data are missing and the contingency table cell counts are relatively small. With missing data in relatively small samples, the likelihood may not have a unique maximum³ under the null and/or the alternative, in which case the score and/or the likelihood ratio statistics will not exist; however, the Bayesian p-values will always exist for a multimodal likelihood. Using one specific Dirichlet prior distribution, the Bayesian p-value is equivalent to the p-value from Fisher's exact test; with 0 cells in the complete case table, for the p-value using the Fisher's exact prior to exist, all marginal counts in the 2×2 table of complete cases must be positive. We note that although Fisher's exact test is commonly used with small cell counts to obtain a conservative p-value by conditioning on both margins of the 2×2 table, we have actually never encountered a study in practice where both margins have been fixed by design.

Unlike the complete case analysis, which may only give correct Type I error when data are missing completely at random (MCAR), our proposed approach yields a valid p-value under the more general case when the data are missing at random (MAR). In our simulations with data that are MAR, we found that the Type I error rates for Bayesian p-values under a range of different Dirichlet prior distributions do not exceed the nominal level. In contrast, under MAR, the complete case p-values for all priors tend to have a Type I error rate that is higher than the nominal level. However, due to the conservativeness of the Fisher's exact p-value, a 'higher Type I error' rate in this context means just a few percentage points above the nominal level. In terms of power, under MAR, the all available data Bayesian p-value based on the Fisher's exact prior tends to have the lowest power. We found that the Bayesian p-value using the prior $\alpha=0.5$ has the highest power (and correct Type I error); thus, we suggest its use with missing data in small samples. Further, even with no missing data, in small samples we suggest use of the Bayesian p-value with $\alpha=0.5$ as an alternative to Fisher's exact test. This Bayesian p-value could be a competitor to use of the so-called mid-exact p-value^{16,17} which has been proposed because of the conservativeness of Fisher's exact test; future research on this topic is warranted. Unfortunately, the mid-p approach does not have a Bayesian formulation (based on a specific prior) and thus the mid-p approach cannot be extended to missing data using the Bayesian approach proposed in this paper. In addition, future research could also examine properties of Bayesian p-values under non-ignorable³ missing data mechanisms. A general discussion of Bayesian approaches to estimation for 2×2 tables with complete data has been proposed by Howard¹⁸, and the approach proposed here with missing data can be applied to extend the methods proposed in that paper.

Acknowledgments

We are grateful for the support provided by grants MH 054693, CA 160679 and CA 06922 from the U.S. National Institutes of Health.

References

1. Lipshultz SE, Wong JC, Lipsitz SR, Simbre VC, Zareba KM, Galpechian V, et al. Frequency of clinically unsuspected myocardial injury at a children's hospital. *Am Heart J*. 2006; 151:916–22. [PubMed: 16569563]
2. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–92.
3. Little RJ, Rubin DB. Statistical analysis with missing data. John Wiley & Sons; 2014.
4. Altham PM. Exact Bayesian analysis of a 2×2 contingency table, and Fisher's "exact" significance test. *J R Stat Soc Ser B Methodol*. 1969; 31:261–9.
5. Altham PM. The analysis of matched proportions. *Biometrika*. 1971; 58:561–76.
6. Casella G, Berger RL. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J Am Stat Assoc*. 1987; 82:106–11.
7. Lin Y, Lipsitz S, Sinha D, Gawande AA, Regenbogen SE, Greenberg CC. Using Bayesian p-values in a 2×2 table of matched pairs with incompletely classified data. *J R Stat Soc Ser C Appl Stat*. 2009; 58:237–46.
8. Chen T, Fienberg SE. Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*. 1974; 30:629–42.
9. Santner TJ, Duffy DE. The statistical analysis of discrete data. Springer Science & Business Media; 2012.
10. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. John Wiley & Sons; 2013.
11. Niederreiter H. Quasi-Monte Carlo methods and pseudo-random numbers. *Bull Am Math Soc*. 1978; 84:957–1041.
12. González J, Tuerlinckx F, De Boeck P, Cools R. Numerical integration in logistic-normal models. *Comput Stat Data Anal*. 2006; 51:1535–48.
13. Chen HY, Little R. A test of missing completely at random for generalised estimating equations with missing data. *Biometrika*. 1999; 86:1–13.
14. Chen T, Fienberg SE. Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*. 1974; 30:629–42.
15. Lipsitz SR, Fitzmaurice GM. The score test for independence in $R \times C$ contingency tables with missing data. *Biometrics*. 1996; 52:751–62. [PubMed: 8672711]
16. Lancaster H. Significance tests in discrete distributions. *J Am Stat Assoc*. 1961; 56:223–34.
17. Routledge R. Practicing safe statistics with the mid-p. *Can J Stat*. 1994; 22:103–10.
18. Howard J. The 2×2 table: A discussion from a Bayesian viewpoint. *Stat Sci*. 1998; 13:351–67.

Table 1

Abnormal levels of myoglobin and diastolic blood pressure for 269 ill children

Myoglobin < 85 ng/ml	Diastolic Blood Pressure < 50 mmHg			
	Yes	No	Missing	Total
Yes	2	0	1	3
No	23	49	144	216
Missing	1	49	0	50
Total	26	98	145	269

Table 2Notation for 2×2 contingency table with missing data

Row Variable	Column Variable			
	1	2	Missing	Total
1	Y_{11}	Y_{12}	Z_{1+}	$Y_{1+} + Z_{1+}$
2	Y_{21}	Y_{22}	Z_{2+}	$Y_{2+} + Z_{2+}$
Missing	U_{+1}	U_{+2}	0	U_{++}
Total	$Y_{+1} + U_{+1}$	$Y_{+2} + U_{+2}$	Z_{++}	n

Table 3

P-values, for the motivating data example in Table 1, for the null hypothesis $H_0: OR = 1$

Alternative	Fisher's exact prior		$\alpha = 0.5$		$\alpha = 1.0$		$\alpha = 1.5$	
	All data	Complete cases	All data	Complete cases	All data	Complete cases	All data	Complete cases
$H_A: OR = 1$	0.095	0.238	0.016	0.051	0.023	0.076	0.014	0.091
$H_A: OR > 1$	0.047	0.118	0.008	0.025	0.012	0.038	0.027	0.046
$H_A: OR < 1$	--- ^a	--- ^a	0.992	0.974	0.988	0.962	0.996	0.954

^aThese posterior probabilities do not exist; the one-sided complete case Fisher's exact P-value for $H_A: OR < 1$ equals 1.

Table 4Joint probabilities for data on (R_1, V_1, R_2, V_2)

Variable 1 (R_1, V_1)	Variable 2 (R_2, Y_2)		
	(0,1)	(0,0)	(1,missing)
(0,1)	$[1 - \pi_1 - \omega_1] \theta_{11}$	$[1 - \pi_1 - \omega_2] \theta_{12}$	$\pi_1 \theta_{1+}$
(0,0)	$[1 - \pi_2 - \omega_1] \theta_{21}$	$[1 - \pi_2 - \omega_2] \theta_{22}$	$\pi_2 \theta_{2+}$
(1,missing)	$\omega_1 \theta_{+1}$	$\omega_2 \theta_{+2}$	0

Table 5

Simulation rejection percentages for two-sided Bayesian p-values for three missing data mechanisms for two-sided 5% significance level test for $H_0: OR = 1$

MAR: $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .45, .05)$											
OR	OR_{CC}	Fisher's exact prior		$\alpha = 0.5$		$\alpha = 1.0$		$\alpha = 1.5$		Complete cases	Complete cases
		All data	Complete cases	All data	Complete cases	All data	Complete cases	All data	Complete cases		
1.00	2.78	2.22	7.05	5.40	19.55	4.50	12.90	3.56	9.30	9.30	9.30
1.55	4.17	3.15	13.55	9.90	31.90	6.70	22.35	5.25	16.70	16.70	16.70
2.33	6.48	7.70	21.15	17.35	45.70	13.45	33.45	11.50	25.81	25.81	25.81
4.00	11.11	14.95	34.95	31.95	62.55	24.20	49.85	19.85	40.30	40.30	40.30
9.00	25.00	28.25	51.95	52.70	79.30	42.25	67.00	35.80	57.20	57.20	57.20

MAR: $(\pi_1, \pi_2, \omega_1, \omega_2) = (.05, .45, .05, .45)$											
OR	OR_{CC}	Fisher's exact prior		$\alpha = 0.5$		$\alpha = 1.0$		$\alpha = 1.5$		Complete cases	Complete cases
		All data	Complete cases	All data	Complete cases	All data	Complete cases	All data	Complete cases		
1.00	0.36	2.15	7.35	5.50	18.45	5.05	12.60	3.65	8.90	8.90	8.90
1.55	0.54	3.10	2.95	8.15	11.10	6.30	6.25	5.27	4.30	4.30	4.30
2.33	0.84	8.75	1.35	18.65	5.35	16.30	3.25	13.80	2.40	2.40	2.40
4.00	1.44	19.20	1.55	34.40	6.90	29.70	5.70	25.70	4.10	4.10	4.10
9.00	3.24	46.85	7.90	67.60	21.20	61.05	16.90	55.40	13.75	13.75	13.75

MCAR: $(\pi_1, \pi_2, \omega_1, \omega_2) = (.25, .25, .25, .25)$											
OR	OR_{CC}	Fisher's exact prior		$\alpha = 0.5$		$\alpha = 1.0$		$\alpha = 1.5$		Complete cases	Complete cases
		All data	Complete cases	All data	Complete cases	All data	Complete cases	All data	Complete cases		
1.00	1.00	2.40	2.35	5.00	5.25	4.15	4.21	3.81	3.85	3.85	3.85
1.55	1.55	3.75	3.50	8.05	8.10	7.15	7.00	6.25	6.10	6.10	6.10
2.33	2.33	10.40	10.35	19.61	19.70	17.60	17.45	15.65	15.80	15.80	15.80
4.00	4.00	26.20	26.15	40.95	40.70	37.20	36.90	34.40	34.00	34.00	34.00
9.00	9.00	52.55	52.00	71.55	71.35	67.70	67.70	63.20	63.50	63.50	63.50