Performing Likelihood Ratio Tests with Multiply-Imputed Data Sets
Author(s): Xiao-Li Meng and Donald B. Rubin
Source: *Biometrika,* Vol. 79, No. 1 (Mar., 1992), pp. 103-111
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2337151
Accessed: 07/03/2011 15:15

Performing likelihood ratio tests with multiply-imputed data sets

By XIAO-LI MENG

Department of Statistics, University of Chicago, Chicago, Illinois 60637, U.S.A.

AND DONALD B. RUBIN

Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

SUMMARY

Existing procedures for obtaining significance levels from multiply-imputed data either (i) require access to the completed-data point estimates and variance-covariance matrices, which may not be available in practice when the dimensionality of the estimand is high, or (ii) directly combine p-values with less satisfactory results. Taking advantage of the well-known relationship between the Wald and log likelihood ratio test statistics, we propose a complete-data log likelihood ratio based procedure. It is shown that, for any number of multiple imputations, the proposed procedure is equivalent in large samples to the existing procedure based on the point estimates and the variance-covariance matrices, yet it only requires the point estimates and evaluations of the complete-data log likelihood ratio statistic as a function of these estimates and the completed data. The proposed procedure, therefore, is especially attractive with highly multiparameter incomplete-data problems since it does not involve the computation of any matrices.

Some key words: Hypothesis testing; Incomplete data; Missing data; p-values; Significance levels; Wald test statistic.

1. INTRODUCTION

Multiple imputation, first proposed by Rubin (1978) and expounded upon by Rubin (1987), is a general statistical technique for handling missing data. It is particularly suited for handling nonresponse in large public-use sample surveys. The key idea, in contrast to single imputation, is to replace each missing value with a set of plausible values, thereby creating multiple completed data sets. Each of these resulting completed data sets is then analyzed using standard complete-data methods. The repeated analyses corresponding to one model for the missing data are combined to create one repeated-imputation inference, which takes proper account of the uncertainty due to missing data under that model. Repeated-imputation inferences across different models are contrasted to reveal sensitivity of inference to posited reasons for the missing data. In this and other ways, multiple imputation retains the major advantages yet rectifies the major disadvantages inherent in single imputation.

Combining the complete-data analyses within a model is usually straightforward. For example, the combined estimate is simply the average of the completed-data estimates. The variability associated with this combined estimate is the sum of two components: the within-imputation variance, i.e. the average of completed-data variance estimates, and the between-imputation component, i.e. the sample variance of the completed-data

estimates. Significance levels can be obtained by computing a modified Wald test statistic from the quantities mentioned above, which is then referred to an F distribution. This procedure has been recently evaluated comprehensively in large samples by Li, Raghunathan & Rubin (1991), with the conclusion that it is essentially calibrated and the loss of power due to a finite number of imputations is quite modest in cases likely to occur in practice. A brief summary of these standard multiple imputation results is presented in § 2.

The procedure studied by Li et al. (1991) requires access to the completed-data estimates and their covariance matrices, and the latter may not be available in practice with standard computer packages, especially when the dimensionality of the estimand is high, such as can occur with partially classified multidimensional contingency tables. Motivated by the well-known relationship between the Wald test statistic and the log likelihood ratio test statistic, we propose a complete-data log likelihood ratio based procedure in § 3. The proposed procedure is equivalent in large samples to the previous one, yet it only requires evaluations of the complete-data log likelihood ratio statistic as a function of parameter estimates for each data set completed by multiple imputation. The computer code for complete-data log likelihood ratio statistics is part of many standard complete-data routines. A numerical example is presented in § 4 to illustrate the implementation of the new procedure. Unless otherwise noted, the appropriate reference for definitions and concepts is Rubin (1987), primarily Chapters 3 and 4.

2. NOTATION AND BACKGROUND

2·1. *Large-sample hypothesis testing with complete data*

Let X be the $n \times p$ complete-data matrix, with associated density $f(X | \psi)$ indexed by the h-dimensional vector parameter ψ. Suppose in a particular analysis the parameter of primary interest, θ, is a k-dimensional vector function of ψ. Let $\hat{\psi} = \hat{\psi}(X)$ be an efficient estimate, e.g. maximum likelihood estimate, of ψ and let $\hat{\theta} = \hat{\theta}(X)$ be an efficient estimate of θ with $U = U(X)$ as the associated covariance matrix. With large samples, a common approximation that is acceptable to both Bayesians and frequentists is

$$U^{-\frac{1}{2}}(\hat{\theta} - \theta) \sim N(0, I), \qquad (2·1)$$

where the sampling distribution of U has lower-order variability than that of $\hat{\theta}$. Consequently, a p-value for the null value of θ, which without loss of generality we take to be zero, can be obtained as $p = \mathrm{pr}(\chi_k^2 > d)$, where

$$d = d(\hat{\theta}, U) = \hat{\theta}^{\mathrm{T}} U^{-1} \hat{\theta} \qquad (2·2)$$

is known as the Wald χ^2 test statistic.

It is well known that, under standard regularity conditions, d is asymptotically equivalent to the log likelihood ratio statistic d', which can be written as

$$d' = d'(\hat{\psi}_0, \hat{\psi} | X) = 2 \log \frac{f(X | \hat{\psi})}{f(X | \hat{\psi}_0)}, \qquad (2·3)$$

where $\hat{\psi}_0 = \hat{\psi}_0(X)$ is an efficient estimate of ψ under $f(X | \psi; \theta = 0)$, that is, with ψ restricted to lie in the $h - k$ dimensional subspace defined by $\theta = 0$. The basis of this equivalence is the h dimensional asymptotic normality of $(\psi - \hat{\psi})$, which we assume in addition to (2·1).

2·2. *Hypothesis testing with multiply-imputed data based on moments*

With incomplete data, write $X = (X_{\text{obs}}, X_{\text{mis}})$, where X_{obs} is the observed part of X and X_{mis} is the missing part. Using proper imputation methods, such as using independent draws from the posterior predictive distribution $f(X_{\text{mis}} | X_{\text{obs}})$, one can create m imputations, $X^{(l)}_{*\text{mis}}$ ($l = 1, \ldots, m$) for the missing values X_{mis}, and thereby create m corresponding completed data sets

$$\{X^{(l)}_*; l = 1, \ldots, m\} = \{(X_{\text{obs}}, X^{(l)}_{*\text{mis}}); l = 1, \ldots, m\}. \qquad (2\cdot4)$$

Standard complete-data methods then can be applied to compute the set of complete-data moments, i.e. estimates and covariance matrices:

$$S_m = \{\hat{\theta}_{*l} \equiv \hat{\theta}(X^{(l)}_*), \; U_{*l} \equiv U(X^{(l)}_*); \; l = 1, \ldots, m\}.$$

The resulting estimate of θ is

$$\bar{\theta}_m = \frac{1}{m} \sum_{l=1}^{m} \hat{\theta}_{*l},$$

with associated variability

$$T_m = \bar{U}_m + (1 + m^{-1}) B_m,$$

where

$$\bar{U}_m = \frac{1}{m} \sum_{l=1}^{m} U_{*l}$$

measures the within imputation variability, and

$$B_m = \frac{1}{m-1} \sum_{l=1}^{m} (\hat{\theta}_{*l} - \bar{\theta}_m)(\hat{\theta}_{*l} - \bar{\theta}_m)^{\text{T}}$$

measures the between imputation variability.

Although hypothesis testing based on S_m can be accomplished directly by forming a Wald type statistic using $\bar{\theta}_m$ and T_m, such a procedure does not provide satisfactory results when the dimensionality of θ is substantially larger than the number of multiple imputations because of the deficient rank of B_m. One way to avoid this problem, which typically occurs in practice, is to assume equal fractions of missing information, that is, to assume that the proportional loss of information is the same for all components of θ. Under this assumption, the Wald type statistic is proportional to

$$D_m = \frac{\bar{\theta}_m^{\text{T}} \bar{U}_m^{-1} \bar{\theta}_m}{k(1 + r_m)}, \qquad (2\cdot5)$$

where

$$r_m = (1 + m^{-1}) \operatorname{tr} (B_m \bar{U}_m^{-1}) / k \qquad (2\cdot6)$$

is the estimated average odds ratio of the fractions of missing information. The exact distribution of D_m is intractable, but it can be approximated well by an F distribution on k and w degrees of freedom, where the best choice of w thus far is given by Li et al. (1991):

$$w = w(r_m) = \begin{cases} 4 + (v-4)\{1 + (1 - \tfrac{1}{2}v) r_m^{-1}\}^2 & \text{if } v = k(m-1) > 4, \\ \tfrac{1}{2}v(1 + 1/k)(1 + r_m^{-1})^2 & \text{otherwise.} \end{cases} \qquad (2\cdot7)$$

Major conclusions of Li et al. (1991) are that, in cases of practical interest, this procedure with $m \geqslant 3$ is insensitive to the assumption of equal fractions of missing information, is well calibrated, and suffers only modest loss of power relative to the ideal likelihood ratio test based directly on X_{obs}.

2·3. Hypothesis testing with multiply-imputed data based on χ^2 statistics

The use of significance levels rather than interval estimates is especially useful and common in models with many parameters. In these cases, however, a standard complete-data analysis often does not provide the entire $k \times k$ variance-covariance matrix $U(X)$, but rather a scalar p-value, or equivalently, the χ^2 statistic, for the null value of θ. Thus, the repeated complete-data computations may not provide the set of the completed-data moments S_m defined in (2·4), but instead, the set of the completed-data Wald χ^2 statistics

$$S_d = \{d_{*l} \equiv d(\hat{\theta}_{*l}, U_{*l}); \, l = 1, \ldots, m\},$$

where the function d is defined in (2·2), or equivalently the set of the completed-data log likelihood ratio statistics

$$S_{d'} = \{d'_{*l} \equiv d'(\hat{\psi}_{0*l}, \hat{\psi}_{*l} | X_*^{(l)}); \, l = 1, \ldots, m\},$$

where the d' function is defined in (2·3), $\hat{\psi}_{0*l} = \hat{\psi}_0(X_*^{(l)})$ and $\hat{\psi}_{*l} = \hat{\psi}(X_*^{(l)})$. Clearly, in these cases, the procedure given by (2·5)–(2·7) cannot be applied directly to obtain significance levels.

A key representation in developing procedures for obtaining significance levels based on S_d or $S_{d'}$ is

$$D_m \overset{\frown}{=} \hat{D}_m \equiv \left(\frac{\bar{d}_m}{k} - \frac{m-1}{m+1} r_m \right) \Big/ (1 + r_m), \qquad (2·8)$$

where $\bar{d}_m$ is the average of $\{d_{*l}; \, l = 1, \ldots, m\}$, and $A \overset{\frown}{=} B$ means $A = B + o(1)$ almost surely. Thus, since $d_{*l} \overset{\frown}{=} d'_{*l}$, asymptotically,

$$D_m \overset{\frown}{=} \hat{D}'_m \equiv \left(\frac{\bar{d}'_m}{k} - \frac{m-1}{m+1} r_m \right) \Big/ (1 + r_m), \qquad (2·9)$$

where $\bar{d}'_m$ is the average of $\{d'_{*l}; \, l = 1, \ldots, m\}$. Notice that the only quantity in (2·8) or (2·9) that is not directly obtainable from S_d or $S_{d'}$ is r_m of (2·6). Replacing r_m with estimates based on S_d or $S_{d'}$ gives a set of procedures for computing p-values when only S_d or $S_{d'}$ is available. The simplest and best replacement so far is given by Li, Meng, Raghunathan & Rubin (1991),

$$\hat{r}_d = (1 + m^{-1}) \left\{ \sum_{l=1}^{m} d_{*l} - \left(\sum_{l=1}^{m} d_{*l}^{\frac{1}{2}} \right)^2 \Big/ m \right\} \Big/ (m-1),$$

with corresponding test statistic

$$\hat{D}_d = \left(\frac{\bar{d}_m}{k} - \frac{m-1}{m+1} \hat{r}_d \right) \Big/ (1 + \hat{r}_d),$$

which is then referred to the distribution F_{k, aw_s}, where $w_s = (m-1)(1 + \hat{r}_d^{-1})^2$ and $a = k^{-3/m}$. The analogue $\hat{D}_{d'}$ replaces d_{*l} with d'_{*l} wherever applicable, and thus $\hat{D}_{d'} \overset{\frown}{=} \hat{D}_d$.

Although the procedures based on S_d or $S_{d'}$ are calibrated reasonably well in many cases of practical importance, they are distinctly inferior to procedures based on S_m, and the loss of power can be severe, especially when k is large and m is small. This degradation in performance is not surprising and will occur for any procedure based only on S_d or $S_{d'}$, because of the extreme loss of information when going from S_m to S_d or $S_{d'}$. Clearly, the only way to overcome this inherent difficulty is to obtain more information. In the next section we show how to combine S_d or $S_{d'}$ with the set of estimates

$$S_e = \{(\hat{\psi}_{*l}, \hat{\psi}_{0*l}); \ l = 1, \dots, m\}$$

to obtain a valid p-value, assuming access to the computer code that calculates the complete-data log likelihood ratio statistic as a function of the parameter estimates at each completed-data set. The new test statistic is not only asymptotically equivalent to the practically satisfactory statistic D_m for any $m \geq 2$, but also can require much less computation because it calculates neither the complete-data variance-covariance matrices U_{*l} $(l = 1, \dots, m)$ nor the between covariance matrix B_m.

3. COMPLETE-DATA LOG LIKELIHOOD RATIO BASED PROCEDURE

3·1. *An intermediate procedure*

Because the numerator of D_m in (2·5) is the simple quadratic form $\bar{d}(\bar{\theta}_m, \bar{U}_m) = \bar{\theta}_m^T \bar{U}_m^{-1} \bar{\theta}_m$, from (2·5), (2·8) and (2·9), we have

$$r_m \simeq \frac{m+1}{k(m-1)} \{\bar{d}_m - d(\bar{\theta}_m, \bar{U}_m)\} \simeq \frac{m+1}{k(m-1)} \{\bar{d}'_m - d(\bar{\theta}_m, \bar{U}_m)\}. \tag{3·1}$$

Since r_m is the only quantity needed for computing the denominator of D_m, we can obtain D_m from the scalar $\bar{d}_m$ or $\bar{d}'_m$ and the scalar distance $d(\bar{\theta}_m, \bar{U}_m)$.

Under the lower order variability assumption of U, $\bar{U}_m \simeq U_{*l}$ for any l. Thus, if we have access to the Wald test statistic $d(\theta_{*l}, U_{*l})$ as a function of θ_{*l} at any completed data set, we can replace $d(\bar{\theta}_m, \bar{U}_m)$ in (3·1) by any $d(\bar{\theta}_m, U_{*l})$, or more efficiently by their average

$$\bar{d}_w = \frac{1}{m} \sum_{l=1}^{m} d(\bar{\theta}_m, U_{*l}). \tag{3·2}$$

The corresponding test statistic then can be constructed as

$$D_w = \frac{\bar{d}_w}{k(1 + r_w)} \simeq D_m, \tag{3·3}$$

where

$$r_w = \frac{m+1}{k(m-1)} (\bar{d}_m - \bar{d}_w). \tag{3·4}$$

3·2. *Motivation for the new procedure*

The calculation of $\bar{d}_w$ in (3·2) requires access to the Wald test statistic in each completed-data set as a function of the estimate of θ, but the well-known asymptotic equivalence between the Wald and log likelihood ratio test statistics suggests that we can compute $d(\bar{\theta}_m, \bar{U}_m)$ assuming access to the function $d'(\hat{\psi}_0, \psi | X)$, defined in (2·3). An important but subtle distinction, however, is that although with the Wald test statistic

we can replace $d(\bar{\theta}_m, \bar{U}_m)$ by any $d(\bar{\theta}_m, U_{*l})$, with the log likelihood ratio test statistic, we must average all m values. That is, even though $d'(\hat{\psi}_{0*l}, \hat{\psi}_{*l} | X_*^{(l)}) \doteq d(\hat{\theta}_{*l}, U_{*l})$ for each l, the analogous replacement of average estimates in d' does not work:

$$d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)}) \neq d(\bar{\theta}_m, U_{*l})$$

for any l, where $\bar{\psi}_{0m}$ and $\bar{\psi}_m$ are the averages of $\{\hat{\psi}_{0*l}; l = 1, \ldots, m\}$ and $\{\hat{\psi}_{*l}; l = 1, \ldots, m\}$, respectively. Rather, asymptotically, with a quadratic complete-data log likelihood function,

$$d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)}) = d(\hat{\theta}_{*l}, U_{*l}) - d(\hat{\theta}_{*l} - \bar{\theta}_m, U_{*l}), \qquad (3\cdot5)$$

which generally does not equal $d(\bar{\theta}_m, U_{*l})$ unless there is no missing information. Their average, however, does equal $\bar{d}_w$ asymptotically because

$$\bar{d}_L \equiv \frac{1}{m} \sum_{l=1}^m d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)}) \doteq d(\bar{\theta}_m, \bar{U}_m). \qquad (3\cdot6)$$

Equation $(3\cdot5)$ follows from viewing the log likelihood ratio $d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)})$ as the log ratio of an h-variate normal density with mean $\hat{\psi}_{*l}$, where the numerator is evaluated at $\bar{\psi}_{0m}$ and the denominator at $\bar{\psi}_m$. The key observation is that, without loss of generality, ψ can be taken such that $\hat{\psi}_{*l} - \hat{\psi}_{0*l} = (\hat{\theta}_{*l}, 0)^T$, where the two components of ψ are orthogonal, and the first component, θ, has covariance U_{*l} and $\hat{\theta}_{0*l} = 0$. The right-hand side of $(3\cdot6)$ follows from $(3\cdot5)$ by viewing $\Sigma\, d(\hat{\theta}_{*l}, U_{*l}) \doteq \Sigma\, d(\hat{\theta}_{*l}, \bar{U}_m)$ as a total sum of squares decomposed into an $m-1$ degrees of freedom sum of squares about the mean, $\Sigma\, d(\hat{\theta}_{*l} - \bar{\theta}_m, \bar{U}_m)$, and a one degree of freedom sum of squares for the mean, $d(\bar{\theta}_m, \bar{U}_m)$.

3·3. Definition and computation of the new procedure

Because $\bar{d}_L \doteq \bar{d}_w$, in view of $(3\cdot3)$ and $(3\cdot4)$, we define our new test statistic as

$$D_L = \frac{\bar{d}_L}{k(1 + r_L)}, \qquad (3\cdot7)$$

where

$$r_L = \frac{m+1}{k(m-1)}(\bar{d}'_m - \bar{d}_L). \qquad (3\cdot8)$$

Since for any $m \geqslant 2$, $D_L \doteq D_m$ and $r_L \doteq r_m$, an obvious approximate reference distribution for D_L is $F_{k, w(r_L)}$, where the $w(r)$ function is defined in $(2\cdot7)$.

To compute D_L, first assume we have S_e and $S_{d'}$, as well as access to the complete-data log likelihood ratio function d' used to compute d'_{*l} $(l = 1, \ldots, m)$. From S_e we calculate the average estimates $\bar{\psi}_{0m}$ and $\bar{\psi}_m$, and from $S_{d'}$ we calculate $\bar{d}'_m$, the average of $\{d'_{*l}; l = 1, \ldots, m\}$. Then, we make a second pass through the completed data, using $\bar{\psi}_{0m}$ and $\bar{\psi}_m$ to calculate the m scalar quantities $d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)})$, and their average, $\bar{d}_L$, defined in $(3\cdot6)$. In most practical applications, the complete-data density $f(X | \psi)$ is from an exponential family so that $d'(\psi_0, \psi | X)$ is linear in the complete-data sufficient statistics, and thus $\bar{d}_L$ is the complete-data log likelihood ratio statistic evaluated at the average parameter estimates $(\bar{\psi}_{0m}, \bar{\psi}_m)$ and at the average completed-data sufficient statistics. Having obtained $\bar{d}'_m$ and $\bar{d}_L$, we apply $(3\cdot8)$ to compute r_L, and then apply

(2·7) and (3·7) to compute $w(r_L)$ and D_L, respectively. Finally, we obtain the significance level $p = \text{pr}\{F_{k,w(r_L)} > D_L\}$.

4. EXAMPLE

4·1. *Likelihood ratio tests for complete contingency tables*

In the analysis of contingency tables, it is very common to test whether the cell probabilities $\{\pi_{ijk...t}\}$ have a special structure corresponding to a parsimonious model. For example, such null models often specify conditional independence of some factors given other factors. With complete data, that is, when each case can be classified into one cell, performing the likelihood ratio test is straightforward. Let $\hat{\pi}_c$ and $\hat{\pi}_{0c}$ be the maximum likelihood estimates of the cell probability for cell c under the saturated and null models, respectively. Then the log likelihood ratio statistic is

$$d' = 2 \sum_c n_c \log (\hat{\pi}_c / \hat{\pi}_{0c}), \qquad (4·1)$$

where the summation is over all the cells c in the table, $n_c = N\hat{\pi}_c$ is the observed count in cell c, and $N = \Sigma_c n_c$ is the total count. When the null model is true, d' is distributed asymptotically as a chi-squared random variable with degrees of freedom equal to the number of independent restrictions on the cell probabilities under the null model (Bishop, Fienberg & Holland, 1975, p. 114).

4·2. *Creating multiple imputations for a partially classified table*

In the presence of missing data, some cases are only partially classified, as illustrated in Table 1 taken from Little & Rubin (1987, Table 9.8). Applying the EM algorithm to find the maximum likelihood estimates for the cell probabilities under the null model and then performing likelihood ratio tests, Little & Rubin (1987, p. 192) considered several models for these data in Table 1 assuming the missing data are missing at random (Rubin, 1976). As an alternative approach, we can use multiple imputation and then apply our new procedure D_L to find the significance level without the need for a covariance matrix for the parameters, as with the moment-based procedure D_m.

The implementation of multiple imputation is particularly simple with a saturated model for a table having a monotone missing data pattern; even for nonmonotone patterns, data augmentation (Tanner & Wong, 1987) makes such multiple imputation

Table 1. *A 2^3 contingency table with partially classified observations*

(*a*) *Completely classified cases*

Clinic, k	Prenatal care, i	Survival, j Died	Survived	
A	Less	3	176	
	More	4	293	
B	Less	17	197	
	More	2	23	$N^{(a)} = 715$ cases

(*b*) *Partially classified cases, clinic missing*

Clinic, k	Prenatal care, i	Survival, j Died	Survived	
?	Less	10	150	
	More	5	90	$N^{(b)} = 255$ cases

straightforward though iterative. For example, for the data set in Table 1, under the noninformative Beta $(1, \ldots, 1)$ prior for the cell probabilities $\{\pi_{ijk}\}$, the lth $(l = 1, \ldots, m)$ imputation of the cell counts can be obtained as follows:

Step 1: Draw the conditional cell probabilities $\pi_{1.ij}$ from Beta $(n_{ij1}^{(a)} + 1, n_{ij2}^{(a)} + 1)$ $(i, j = 1, 2)$, where $n_{ijk}^{(a)}$ is the observed cell count in cell (i, j, k) of the completely classified Table 1(a).

Step 2: Given the conditional probabilities $\pi_{1.ij}$ $(i, j = 1, 2)$ drawn in Step 1, draw $n_{ij1}^{(b)}$ from Binomial $(n_{ij+}^{(b)}, \pi_{1.ij})$, where $n_{ij+}^{(b)}$ is the observed counts in the marginal cell (i, j) of the partially classified Table 1(b); let $n_{ij2}^{(b)} = n_{ij+}^{(b)} - n_{ij1}^{(b)}$.

Step 3: Calculate the lth imputed cell count for cell (i, j, k) as $n_{ijk}^{(l)} = n_{ijk}^{(a)} + n_{ijk}^{(b)}$ $(i, j, k = 1, 2)$.

4·3. *Testing null models using our new procedure*

Once these m completed tables are created, one can treat each as if it were the complete data, and use formula (4·1) to compute m completed-data log likelihood ratio test statistics

$$d'_{*l} = 2 \sum_c n_c^{(l)} \log (\hat{\pi}_c^{(l)} / \hat{\pi}_{0c}^{(l)}) \quad (l = 1, \ldots, m).$$

The computation of $\bar{d}_L$ of (3·6) involves evaluating the log likelihood ratio function at $\bar{\psi}_m$ and $\bar{\psi}_{0m}$, where ψ is a vector parameter providing the cell probabilities. For simplicity, we take ψ to be the cell probabilities, whence $\bar{\psi}_m$ and $\bar{\psi}_{0m}$ are simply the averages of cell probabilities under the saturated and null models, respectively, across the imputations. Thus, letting $\bar{\pi}_c$ and $\bar{\pi}_{0c}$ be the averages of the estimated cell probabilities across imputations for cell c under the saturated and null models, respectively, by (4·1),

$$d'(\bar{\psi}_{0m}, \bar{\psi}_m | X_*^{(l)}) = 2 \sum_c n_c^{(l)} \log (\bar{\pi}_c / \bar{\pi}_{0c}) \quad (l = 1, \ldots, m),$$

whose average is equal to

$$\bar{d}_L = 2 \sum_c \bar{n}_c \log (\bar{\pi}_c / \bar{\pi}_{0c}) = 2 \sum_c \bar{n}_c \log (\bar{n}_c / \bar{n}_{0c}),$$

where $\bar{n}_c = N\bar{\pi}_c$ and $\bar{n}_{0c} = N\bar{\pi}_{0c}$.

Table 2 presents the results for testing two different null models with several choices of m. The conditional independence model asserts that conditional on clinic, survival and prenatal care are independent, which was Little & Rubin's (1987) preferred parsimonious model based on their tests. The full independence model has all three factors mutually independent, and is included here for comparison. The second column gives r_L of (3·8), the estimated average odds ratio of the fractions of missing information,

Table 2. *Testing null models using D_L*

	Conditional independence			Full independence		
m	r_L	D_L	p	r_L	D_L	p
2	0·46	0·06	0·938	0·70	41·86	0·000
3	0·32	0·08	0·927	0·41	51·12	0·000
5	0·28	0·02	0·980	0·26	56·36	0·000
7	0·26	0·03	0·966	0·29	54·35	0·000
10	0·30	0·08	0·928	0·34	52·33	0·000
25	0·32	0·10	0·905	0·36	51·97	0·000
50	0·30	0·09	0·919	0·37	50·98	0·000
Ideal		0·10	0·904		52·67	0·000

which, except for $m = 2$, is relatively stable as m increases, as we expected. The third column gives the value of the new test statistic D_L, and the fourth column gives the corresponding p-value using the F reference distribution. The last three columns are the counterparts for the full independence model. For comparison, we include in the last row the corresponding results using the ideal likelihood ratio test based directly on the data in Table 1. It is clear from Table 2 that the tests based on D_L reach essentially the same practical conclusions as the ideal test, even for small m. The satisfactory performance of these tests based on small numbers of multiple imputations is especially encouraging, because small m is common in practice.

REFERENCES

BISHOP, Y. M. M., FIENBERG, S. E. & HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

LI, K. H., MENG, X. L., RAGHUNATHAN, T. E. & RUBIN, D. B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* **1**, 65-92.

LI, K. H., RAGHUNATHAN, T. E. & RUBIN, D. B. (1991). Large sample significance levels from multiply-imputed data using moment-based statistics and an F reference distribution. *J. Am. Statist. Assoc.* **86**, 1065-73.

LITTLE, R. J. A. & RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-92.

RUBIN, D. B. (1978). Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. In *Proc. Survey Res. Meth. Sect., Am. Statist. Assoc.*, pp. 20-34. Also in *Imputation and Editing of Faulty or Missing Survey Data*, pp. 1-23. U.S. Dept. of Commerce, Bureau of the Census.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys*. New York: Wiley.

TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.* **82**, 805-11.