STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:4086–4093
Published online 12 March 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3239

An imputation strategy for incomplete longitudinal ordinal data

Hakan Demirtas*,† and Donald Hedeker

Division of Epidemiology and Biostatistics, University of Illinois at Chicago, Chicago, IL 60612, U.S.A.

SUMMARY

A new quasi-imputation strategy for correlated ordinal responses is proposed by borrowing ideas from random number generation. The essential idea is collapsing ordinal levels to binary ones and converting correlated binary outcomes to multivariate normal outcomes in a sensible way so that re-conversion to the binary and then ordinal scale, after conducting multiple imputation, yields the original marginal distributions and correlations. This conversion process ensures that the correlations are transformed reasonably, which in turn allows us to take advantage of well-developed imputation techniques for Gaussian outcomes. We use the phrase 'quasi' because the original observations are not guaranteed to be preserved. We present an application using a data set from psychiatric research. We conclude that the proposed method may be a promising tool for handling incomplete longitudinal or clustered ordinal outcomes. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: longitudinal data; missing data; multiple imputation; random number generation

1. INTRODUCTION

Missing data are common in longitudinal studies. Determining a suitable analytical approach in the presence of incomplete observations is a major focus of scientific inquiry due to the additional complexity that arises through missing data. Incompleteness generally complicates the statistical analysis in terms of biased parameter estimates, reduced statistical power, and degraded confidence intervals, leading to false inferences [1].

Advances in computational statistics have produced flexible missing-data procedures with a sound statistical basis. One of these procedures involves multiple imputation (MI) [2, 3], a simulation technique that replaces each missing datum with a set of $m>1$ plausible values. The m versions of complete data are then analyzed by standard complete-data methods and the results are combined into a single inferential statement using arithmetic rules to yield estimates, standard

*Correspondence to: Hakan Demirtas, Division of Epidemiology and Biostatistics (MC 923), University of Illinois at Chicago, 1603 West Taylor Street, Chicago, IL 60612, U.S.A.
†E-mail: demirtas@uic.edu

Received 16 January 2008
Accepted 16 January 2008

errors, and p-values that formally incorporate missing-data uncertainty to the modeling process. The key ideas and advantages of MI were reviewed by Rubin [4] and Schafer [5].

Imputation strategies for longitudinal or clustered continuous responses [5–12] and for correlated binary outcomes have been an active area of research [5, 13–15]. In this article, we extend the imputation methodology for correlated binary data developed by Demirtas and Hedeker [16] to an ordinal data setting, borrowing ideas from random number generation. Specifically, generation of multivariate Gaussian outcomes, given the marginal distribution and correlation structure of the binary data that are obtained through collapsing ordinal levels, taking advantage of distributional features of normality in the imputation process, is presented. A central methodological focus of this article is MI under a multivariate normality assumption with re-conversion to the ordinal scale while preserving key distributional properties.

We use the phrase 'quasi-imputation' because the original values of the observed ordinal outcomes are not guaranteed to be preserved. In this regard, we argue that if the goal is creating several imperfect proxies of augmented version of the data sets in order to account for uncertainty due to missing data in addition to ordinary sampling variability, it is not necessary to strictly adhere to the established definition of MI. Furthermore, for the purpose of this paper, the focus is limited to the analysis of a single ordinal variable over time. Applications to multivariate repeated measures are computationally more challenging, but the fundamental issues are the same. The ideas presented are equally applicable to other forms of correlated ordinal data (e.g. clustered data).

The organization of this paper is as follows: In Section 2, we focus on imputing correlated binary outcomes and re-visit the method of Demirtas and Hedeker [16]. In Section 3, we present an extension that is designed for incomplete ordinal data. In Section 4, we apply the proposed methodology to a real data set from psychiatric research. Section 5 includes concluding remarks and discussion.

2. IMPUTING BINARY DATA

A refresher on binary variate generation is needed for describing the operational characteristics of imputing binary data. Emrich and Piedmonte [17] proposed a random number generation algorithm for binary data. Let $Y_1, \ldots, Y_j$ represent binary variables such that $E[Y_j] = p_j$ and $\text{Corr}(Y_j, Y_k) = \delta_{jk}$, where p_j $(j = 1, \ldots, J)$ and δ_{jk} $(j = 1, \ldots, J-1, k = 2, \ldots, J)$ are given, where $J \geqslant 2$. As Emrich and Piedmonte [17] noted, δ_{jk} is bounded below by $\max(-\sqrt{(p_j p_k / q_j q_k)}, -\sqrt{(q_j q_k / p_j p_k)})$ and above by $\min(\sqrt{(p_j q_k / q_j p_k)}, \sqrt{(q_j p_k / p_j q_k)})$, where $q = 1 - p$. Let $\Phi[x_1, x_2, \rho]$ be the cumulative distribution function for a standard bivariate normal random variable with correlation coefficient ρ. Naturally, $\Phi[x_1, x_2, \rho] = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(z_1, z_2, \rho)\, dz_1\, dz_2$, where $f(z_1, z_2, \rho) = [2\pi(1-\rho^2)^{1/2}]^{-1} \times \exp[-(z_1^2 - 2\rho z_1 z_2 + z_2^2)/(2(1-\rho^2))]$. We could generate multivariate normal outcomes (Z's) by solving the following equation:

$$\Phi[z(p_j), z(p_k), \rho_{jk}] = \delta_{jk}(p_j q_j p_k q_k)^{1/2} + p_j p_k \tag{1}$$

for ρ_{jk} $(j = 1, \ldots, J-1, k = 2, \ldots, J)$, where $z(p)$ denotes the pth quantile of the standard normal distribution. As long as δ_{jk} satisfies the range condition mentioned above, the solution is unique. Repeating this numerical integration process $J(J-1)/2$ times, one can obtain the overall correlation matrix (say Σ) for the J-variate standard normal distribution with mean 0.

DOI: 10.1002/sim

However, it should be noted that positive definiteness of Σ cannot be guaranteed. To create dichotomous outcomes (Y_j) from the generated normal outcomes (Z_j), we set $Y_j = 1$ if $Z_j \leqslant z(p_j)$ and 0 otherwise for $j = 1, \ldots, J$. This produces a vector with the desired properties: $E[Y_j] = P(Y_j = 1) = P(Z_j \leqslant z(p_j)) = p_j$. $\mathrm{Cov}(Y_j, Y_k) = P(Y_j = 1, Y_k = 1) - p_j p_k = P(Z_j \leqslant z(p_j), Z_k \leqslant z(p_k)) - p_j p_k = \Phi[z(p_j), z(p_k), \rho_{jk}] - p_j p_k = \delta_{jk}(p_j q_j p_k q_k)^{1/2}$. Therefore, $\mathrm{Corr}(Y_j, Y_k) = \mathrm{Cov}(Y_j, Y_k)/(p_j q_j p_k q_k)^{1/2} = \delta_{jk}$ by equation (1). Emrich and Piedmonte's [17] method is essentially based on the idea of borrowing information from the higher-order moments of the multivariate normal distribution to generate binary variates. Demirtas and Hedeker [16] use the first step of their algorithm, i.e. converting the correlated binary outcomes to multivariate normal outcomes in a sensible way so that re-conversion to the binary scale, after performing MI, yields the original specified marginal expectations and correlations. This conversion process ensures that the correlations are transformed reasonably, which in turn allows researchers to take advantage of well-developed imputation techniques for Gaussian outcomes [5, 15].

Demirtas and Hedeker [16] articulate on how this paradigm can be used to create multiply imputed data sets. To set the notation, we assume that there are N subjects and the ith subject is to be observed at J occasions, $t_{i1}, \ldots, t_{iJ}$. We denote the complete response vector for subject i as $Y_i^{\mathrm{c}} = (Y_{i1}, \ldots, Y_{iJ})^{\mathrm{T}}$. Stacking the $N Y_i^{\mathrm{c}}$ vectors yields the overall $N \times J$ response matrix $\mathscr{Y}$ whose columns correspond to the variables Y_j (where $j = 1, \ldots, J$). To allow for missing data, we introduce the indicator variables $R_i = (R_{i1}, \ldots, R_{iJ})^{\mathrm{T}}$, where $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ otherwise. Owing to drop-out or intermittent missingness, the ith subject is observed at n_i occasions, with $n_i < J$. Let Y_i^{obs} denote the $n_i \times 1$ vector of observed responses on the ith individual. That is, $Y_i^{\mathrm{obs}} = \Delta_i Y_i^{\mathrm{c}}$ is the observed portion of Y_i^{c}, where Δ_i is an $n_i \times J$ matrix, comprised of non-zero rows of $\mathrm{diag}(R_{i1}, \ldots, R_{iJ})$. Finally, let $\mathscr{Y} = (\mathscr{Y}_{\mathrm{obs}}, \mathscr{Y}_{\mathrm{mis}})$ where $\mathscr{Y}_{\mathrm{obs}}$ and $\mathscr{Y}_{\mathrm{mis}}$ denote the observed and missing parts of the complete-data matrix $\mathscr{Y}$, respectively.

The steps of the quasi-imputation strategy for binary data are as follows.

1. Given an $N \times J$ matrix of incomplete binary responses, find the marginal expectations, $E[Y_j] = p_j$ and Pearson correlations $\mathrm{Corr}(Y_j, Y_k) = \delta_{jk}$ based on the observed data, $\mathscr{Y}_{\mathrm{obs}}$. Form the $N \times J$ missingness indicator matrix R, consisting of the R_{ij} elements.
2. Find normal correlations, ρ_{jk}, by solving the numerical integration in equation (1) for each element in the $J \times J$ correlation matrix (i.e. $J(J-1)/2$ times). Obtain the Cholesky square root (Σ^{chol}) of the resulting correlation matrix Σ.
3. For each partly or completely observed binary response variable Y_j, generate standard univariate normal variates, Z_j's, of length $2N$. Store the non-zero elements of $Z_{1j} = Z_j I[Z_j > \Phi^{-1}(1-p_j)]$ and $Z_{0j} = Z_j I[Z_j < \Phi^{-1}(1-p_j)]$, where I is the indicator function. Draw random samples from Z_{1j} and Z_{0j} without replacement for $Y_j = 1$ and 0, respectively. Sizes of these random samples are determined by the lengths of ($Y_j = 1$) and ($Y_j = 0$). Denote the new continuous response matrix as $\mathscr{Z}$. When Y_{ij}'s are missing ($R_{ij} = 0$), assign a 0 to $\mathscr{Z}_{ij}$. Note that the original binary data matrix ($\mathscr{Y}$) and the newly formed continuous data matrix are of the same dimension ($N \times J$).
4. Right multiply $\mathscr{Z}$ by Σ^{chol} to obtain the Gaussian data matrix $\mathscr{Y}^{\mathrm{new}}$ whose elements are ${\mathscr{Y}_{ij}}^{\mathrm{new}}$.
5. Assign missing values in $\mathscr{Y}^{\mathrm{new}}$ using R to reflect the original non-response structure.
6. Perform MI assuming multivariate normality. This can be done by first running an EM-type algorithm [18], and then employing a data augmentation procedure [19], as implemented in some software packages (e.g. Splus). The EM algorithm is useful for two reasons: it provides

DOI: 10.1002/sim

excellent starting values for the data augmentation scheme, and it gives us an idea about the convergence behavior. Data augmentation using the Bayesian paradigm has been perceived as a natural tool to create multiply imputed data sets. For further details, see Schafer [5] and Schimert *et al.* [15].

7. Convert the numbers back to the binary scale by using the initial marginal expectations as cutoff points according to a univariate normal cumulative distribution function. In other words, Y_{ij} takes the value 1 if $P(\mathscr{Y}_{ij}{}^{\text{new}} \leqslant z(p_j))$ and 0 otherwise.
8. Conduct any appropriate statistical analyses; find the point estimates and standard errors for the parameters of interest and combine them using Rubin's [3] rules.

If there are incompletely or completely observed binary covariates, we can include them in the Gaussianization process. The rationale for this is that it may reduce the fraction of missing information for the parameters under consideration in the subsequent analyses. Similarly, if there are fully observed continuous covariates, we can use them in the Gaussian imputation model, reflecting a belief that doing so may recover some of the missing information to the extent that they are the correlates or causes of missingness and/or responses (see Demirtas [20]). Furthermore, the reason that the length of Z_j's $(2N)$ is chosen to be larger than the number of subjects (N) in Step 3 is to allow for a sufficient length of numbers in the random selection process when forming the new data matrix. The rationale of setting $\mathscr{Z}_{ij}=0$ when $R_{ij}=0$ in Step 3 is to nullify the effect of missing value locations in the subsequent right multiplication of $\mathscr{Z}_{ij}$ with Σ^{chol}.

In the following section, we extend this methodology to an ordinal data setting.

3. IMPUTING CORRELATED ORDINAL DATA

Demirtas [21] proposed an iterative algorithm for generating correlated ordinal variables with partial specification of first- and second-order moments, which we elaborate on below.

The problem can be stated as generating J ordinal outcomes, $Y_1, Y_2, \ldots, Y_J$ such that $P(Y_j=k)=p_{jk}$ and $\text{corr}(Y_i, Y_j)=\delta_{ij}^{\text{ORD}}$, where p_{jk}, $j=1,2,\ldots,J$, $k=1,2,\ldots,K$, and δ_{ij}^{ORD}, $i=1,\ldots,J-1$, $j=2,\ldots,J$, are specified for $J\geqslant 2$ and $K\geqslant 3$. δ_{ij}^{ORD} and p_{jk} are assumed to be independent. Steps of the approach in Demirtas [21] are as follows:

1. Collapse the ordinal categories to binary ones. If the number of categories K is even, assign 0 when $Y_j=1,2,\ldots,K/2$ and 1 when $Y_j=K/2+1,\ldots,K$. If K is odd, the first $(K-1)/2$ categories become 0 and the last $(K-1)/2$ categories become 1. Median category $(K+1)/2$ is assigned to a suitable binary category that makes the expectation closest to 0.5 since binary variables behaves best in the neighborhood of 0.5. Denote the newly formed binary variables as Y_j^{BIN} with $p_j^{\text{BIN}}=E[Y_j^{\text{BIN}}]$. Under this scheme, marginal expectations become $p_j^{\text{BIN}}=\sum_{k=K/2+1}^{K} p_{jk}$ for even K, and $p_j^{\text{BIN}}=\sum_{k=(K+3)/2}^{K} p_{jk}$ or $p_j^{\text{BIN}}=\sum_{k=(K+1)/2}^{K} p_{jk}$ for odd K, depending on the classification of the median category.
2. Find corresponding binary correlations δ_{ij}^{BIN} via simulation. Asymptotically, $|\delta_{ij}^{\text{BIN}}|\geqslant|\delta_{ij}^{\text{ORD}}|$ for all i and j as $N\rightarrow\infty$. The reason is that when collapsing is done, the product of standard errors in the denominator decreases at a faster rate than the covariance term in the numerator in large samples. Then, take a large number of replicates N such as 100 000 and generate binary variables with p_j^{BIN} and $\delta_{ij}^{\text{BIN}}=\delta_{ij}^{\text{ORD}}$. After generating the $N\times J$ matrix

DOI: 10.1002/sim

of binary data, go back to the ordinal scale using the original proportions $p_{jk}/(1-p_j^{\mathrm{BIN}})$ and $p_{jk}/p_j^{\mathrm{BIN}}$ for binary categories 0 and 1, respectively, absolute ordinal correlations will be underestimated. Set up an iterative algorithm in which $|\delta_{ij}^{\mathrm{BIN}}|$ is increased until $\delta_{ij}^{\mathrm{ORD}}$ converges to the true value after ordinal-binary-ordinal conversion. This optimization process should be implemented for each pair of variables (i.e. the process is repeated $J(J-1)/2$ times), leading to an overall binary correlation matrix Σ whose off-diagonal elements are $\delta_{ij}^{\mathrm{BIN}}$.
3. Once proper binary correlations (elements of Σ) are found, generate ordinal random variables by first simulating binary data, then perform the same conversion process in Step 2 using the original proportions $p_{jk}/(1-p_j^{\mathrm{BIN}})$ and $p_{jk}/p_j^{\mathrm{BIN}}$ for any desired number of replicates (sample size).

Our proposed MI approach is based on collapsing the ordinal levels to binary ones and finding an appropriate correlation structure of dichotomized data utilizing the first two steps in Demirtas [21] and inflating the binary correlations on purpose (rationale appears at the last stage). The next stage is employing the first seven steps in Demirtas and Hedeker [16]. The final stage is going back to the ordinal scale by preserving the relative sample sizes in each category with a similar spirit in Step 3 of Demirtas [21]. It is a relatively straightforward extension of binary data imputation. Specifically, algorithm is as follows:

1. Find observed marginal values and correlations ($\delta_{jk}^{\mathrm{ORD}}$) for the ordinal data.
2. Dichotomize the data. If the number of distinct ordinal categories is even, it is easy. If it is odd, find the best combination (the one that makes marginal expectations closest to 0.5, since binary variables behave best near to 0.5).
3. Find observed binary correlations ($\delta_{jk}^{\mathrm{BIN}}$) via iterative simulation (Step 2 in Demirtas [21]).
4. In equation (1), solve for ρ_{jk} using marginal means for dichotomized data and correlations ($\delta_{jk}=\delta_{jk}^{\mathrm{BIN}}(\delta_{jk}^{\mathrm{BIN}}/\delta_{jk}^{\mathrm{ORD}})$) and then generate multivariate normals. When we move from ordinal scale to binary scale, absolute correlations are going to increase. Hence, δ_{ij}'s in equation (1) will be overestimated (in absolute value sense) on purpose.
5. Generate multivariate normal variates as in binary case and impose missing values as in the original data set.
6. Proceed with well-known imputation techniques for continuous data [5].
7. After multiply imputing incomplete-data sets, go back to binary scale.
8. Finally, go back to ordinal scale by preserving the relative sample sizes in each category. When we do that, overestimated correlations will become what they should be (similar to Step 3 in Demirtas [21]).

4. APPLICATION

Our data example comes from the National Institute of Mental Health Schizophrenia Collaborative Study [22]. We examined Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS). Item 79, 'Severity of Illness' was scores as 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; 7 = extremely ill. We re-coded the seven ordered categories into four as Hedeker and Gibbons [22] did: (1) normal or borderline mentally ill, (2) mildly or moderately ill, (3) markedly ill, and (4) severely or extremely ill. In this study, patients were randomly assigned to receive either an anti-psychotic drug or placebo.

DOI: 10.1002/sim

Measurements were planned for weeks 0, 1, 3, and 6, but missing values occurred primarily due to drop-out. A few subjects had missing measurements and subsequently returned, for simplicity, we have removed these. A small number of measurements were also taken at intermediate time points (weeks 2, 4, and 5) which we also ignore. With these exclusions, the sample contains 312 patients who received a drug and 101 who received a placebo. There is a fair amount of attrition; about 25 per cent of people did not complete the study. In the drug group, 3 patients dropped out immediately after week 0, 27 dropped out after week 1, 34 dropped out after week 3, and 248 completed the study. In the placebo group, no patients dropped out after week 0, 18 dropped out after week 1, 19 dropped out after week 3, and there were 64 completers. Hedeker and Gibbons [22] noted that the mean response profiles are approximately linear when plotted against the square root of week, and they express time on the square-root scale in their models.

We implemented the proposed methodology on this psychiatric data set. First, we dichotomized the ordinal levels. Then, inflated binary correlations were obtained through an iterative procedure as outlined in Section 3. Although the treatment indicator is not a part of outcomes, we included this variable when we created multivariate normal outcomes since previous analysis showed that it has substantial impact on inferences. After creating 10 multiply imputed data sets, we re-transformed continuous data to the binary scale, followed by the binary–ordinal conversion.

We analyzed the resulting data sets by a random intercept and slope mixed-effects ordinal regression model. For subject i at timepoint j, for $c-1$ cumulative logits (here, $c=4$), with D denoting Drug (0 for placebo, 1 for anti-psychotic drug) and W denoting Week,

$$\log\left[\frac{P_{ijc}}{1-P_{ijc}}\right]=\gamma_c-[\beta_0+\beta_1\sqrt{W_j}+\beta_2 D_i+\beta_3(D_i\times\sqrt{W_j})+v_{0i}+v_{1i}\sqrt{W_j}]$$

where β's stand for fixed effects, v_{0i} is the random intercept, v_{1i} is the random slope, and γ's stand for thresholds ($\gamma_1=0$). Random effects are assumed to follow a normal distribution. In this model, $-\beta_0$ represents the week 0 first logit (category 1 *versus* 2–4), $\gamma_1-\beta_0$ the week 0 second logit (1–2 *versus* 3–4), and $\gamma_2-\beta_0$ the week 0 third logit (1–3 *versus* 4) for the placebo group. In terms of the regression parameters, β_1 represents the weekly (in square root units) logit change for placebo patients, β_2 is the difference in the week 0 for drug patients, and β_3 is the difference in the weekly (square root) logit change between drug and placebo groups. The random subject effects v_{0i} and v_{1i} represent intercept and slope deviations for subject i, respectively. The estimated coefficients for fixed effects, their standard errors, z- and p-values are shown in Table I. The results indicate that the treatment groups do not significantly differ at baseline, the placebo group improves over time, and that the drug group has greater improvement over time, compared with the placebo group. These conclusions are consistent with the previously published findings [10, 22].

Table I. Random intercept and slope ordinal logistic regression results for fixed effects.

Parameter	Estimate	Std. error	z-Value	p-Value
β_0	6.291	0.386	16.30	<0.0001
β_1	−0.762	0.179	−4.26	<0.0001
β_2	0.184	0.391	0.47	0.32
β_3	−1.182	0.202	−5.85	<0.0001
γ_1	2.793	0.226	12.36	<0.0001
γ_2	5.112	0.299	17.09	<0.0001

DOI: 10.1002/sim

5. DISCUSSION

A major imputation principle is not to distort the marginal distributions and associations between observed and imputed variables. In an attempt to follow this principle, we proposed a quasi-imputation strategy for correlated ordinal responses. The core idea is reasonably converting ordinal variables to Gaussian ones, then creating multiply imputed data sets by well-known imputation procedures that assume multivariate normality, and finally re-converting Gaussian outcomes to the original ordinal scale with dichotomization as an intermediate stage.

There are a few limitations, however. First of all, the variance–covariance structure is not usually independent of the means for categorical variables. As we explain in Section 2, binary correlations have upper and lower limits that are imposed by their marginal means. If the range condition is not satisfied, the solution for the normal correlations is not unique and the method is not safely applicable. Second, even when the range assumption is not violated, the overall normal correlation matrix that is formed through pairwise correlations (which are obtained by solving a series of numerical integration problems) is not guaranteed to be positive definite.

The approach we take in this paper focuses on MI, whose key advantages are well documented [5]. When a direct maximum likelihood procedure is available for a particular analysis, it may indeed be the convenient method. However, MI still offers some unique advantages for data analysts. First, MI allows researchers to use more conventional models and software; an imputed data set may be analyzed by literally any method that would be suitable if the data were complete. As computing environments and statistical models grow increasingly sophisticated, the value of using familiar methods and software becomes important. Second, there are still many classes of problems for which no direct maximum likelihood procedure is available. Even when such a procedure exists, MI can be more attractive due to fact that the separation of the imputation phase from the analysis phase lends greater flexibility to the entire process. Lastly, MI singles out missing data as a source of random variation distinct from ordinary sampling variability.

Given that forcing incomplete-data sets to a full rectangular format through MI has long been recognized as a practically useful approach in medical statistics, the proposed methodology could be a good addition to the researchers' methodological toolkit when dealing with incomplete correlated ordinal responses. Although it is not applicable in every scenario, we believe that this easy-to-implement, conceptually simple approach is noteworthy in the sense that it preserves the marginal distributions and associations between the observed and imputed variables through a novel ordinal-binary-Gaussian-ordinal-binary conversion scheme. The method is currently designed to operate under ignorable missingness mechanisms and can be regarded as a building block for the extensions that may potentially accommodate non-ignorable non-response.

REFERENCES

1. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: New York, 2002.
2. Rubin DB. Formalizing subjective notions about the effect of non-respondents in sample surveys. *Journal of the American Statistical Association* 1977; **72**:538–543.
3. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
4. Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association* 1996; **91**:473–520.
5. Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall: London, 1997.
6. Schafer JL. *PAN*: *Multiple Imputation for Multivariate Panel Data*, *Software Library for S-PLUS*. Department of Statistics, The Pennsylvania State University: University Park, PA, 1997.

DOI: 10.1002/sim

7. Schafer JL. Multiple imputation with PAN. In *New Methods for the Analysis of Change 355–377*, Sayer AG, Collins LM (eds). American Psychological Association: Washington, DC, 2001.
8. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002; **11**:437–457.
9. Liu M, Taylor JM, Belin TR. Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics* 2000; **56**:1157–1163.
10. Demirtas H, Schafer JL. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2003; **22**:2553–2575.
11. Demirtas H. Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2005; **24**:2345–2363.
12. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
13. van Buuren S, van Rijckevorsel JLA. Imputation of missing categorical data by maximizing internal consistency. *Psychometrika* 1992; **57**:567–580.
14. van Buuren S, Oudshoom CGM. *MICE V1.0 Users Guide*. TNO Preventie en Gezonheid, Leiden, 2000; TNO/PG/V GZ 00.038.
15. Schimert J, Schafer JL, Hesterberg T, Fraley C, Clarkson DB. *Analyzing Data with Missing Values in S-Plus*. Data Analysis Products Division, Insightful Corp.: Seattle, WA, 2001.
16. Demirtas H, Hedeker D. Gaussianization-based quasi-imputation and expansion strategies for incomplete correlated binary responses. *Statistics in Medicine* 2007; **26**:782–799.
17. Emrich LJ, Piedmonte RP. A method for generating high-dimensional multivariate binary outcomes. *American Statistician* 1991; **45**:302–304.
18. Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of Royal Statistical Society*, *Series B* 1977; **39**:1–38.
19. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association* 1987; **82**:528–540.
20. Demirtas H. Simulation-driven inferences for multiply imputed longitudinal datasets. *Statistica Neerlandica* 2004; **58**:466–482.
21. Demirtas H. A method for multivariate ordinal data generation given marginal distributions and correlations. *Journal of Statistical Computation and Simulation* 2006; **76**:1017–1025.
22. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994; **50**:933–944.

DOI: 10.1002/sim