Received: 15 February 2017 | Revised: 8 November 2017 | Accepted: 12 November 2017

DOI: 10.1002/sim.7583

RESEARCH ARTICLE

WILEY Statistics in Medicine

Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout

Yongqiang Tang

Shire, 300 Shire Way, Lexington, MA 02421, USA

Correspondence
Yongqiang Tang, Shire, 300 Shire Way, Lexington, MA 02421, USA.
Email: yongqiang_tang@yahoo.com

The controlled imputation method refers to a class of pattern mixture models that have been commonly used as sensitivity analyses of longitudinal clinical trials with nonignorable dropout in recent years. These pattern mixture models assume that participants in the experimental arm after dropout have similar response profiles to the control participants or have worse outcomes than otherwise similar participants who remain on the experimental treatment. In spite of its popularity, the controlled imputation has not been formally developed for longitudinal binary and ordinal outcomes partially due to the lack of a natural multivariate distribution for such endpoints. In this paper, we propose 2 approaches for implementing the controlled imputation for binary and ordinal data based respectively on the sequential logistic regression and the multivariate probit model. Efficient Markov chain Monte Carlo algorithms are developed for missing data imputation by using the monotone data augmentation technique for the sequential logistic regression and a parameter-expanded monotone data augmentation scheme for the multivariate probit model. We assess the performance of the proposed procedures by simulation and the analysis of a schizophrenia clinical trial and compare them with the fully conditional specification, last observation carried forward, and baseline observation carried forward imputation methods.

KEYWORDS

control-based pattern mixture model, delta-adjusted pattern mixture model, fully conditional specification, missing not at random, monotone data augmentation, tipping point analysis

1 | INTRODUCTION

Missing data are unavoidable, and they represent a potential source of bias in a clinical trial.[1] The primary analysis generally assumes that the data are missing at random (MAR), or equivalently that the probability of missingness may depend on the observed data, but is unrelated to the missing outcomes after conditioning on the observed data.[2] The MAR mechanism implies that after dropout, participants will have the same future statistical behavior as those from the same treatment group who have identical history but continue the treatment.[2-4] Under MAR, all participants randomized to the experimental arm are expected to receive the same amount of benefit at the end of trial no matter how long they stay on the treatment. The MAR assumption may not be appropriate particularly for the trial investigating a symptomatic

treatment, in which the symptoms of the disease may return, and the treatment benefit is not expected to persist after the treatment discontinuation irrespective of the reasons for discontinuation.

The CHMP guideline[5] and a recent full data augmentation (FDA)–mandated panel report from the National Research Council[6] recommend sensitivity analysis under the missing not at random (MNAR) mechanism, under which, the probability of missingness is related to the unobserved outcome after controlling for the observed data. An example would be that a participant who quit the trial because he becomes too ill to come to the clinic. Under MNAR, there is a systematic difference between dropouts and completers, and such difference cannot be completely explained by the observed data.

The controlled imputation methods including the control-based[7,8] and delta-adjusted[9,10] pattern mixture models (PMMs) have been popularly used as MNAR sensitivity analyses in new drug applications in recent years. The PMMs make clinically plausible and interpretable assumptions about the missing outcomes after dropout. The response profile after participants in the experimental arm discontinue the treatment is assumed to become worse compared to participants who remain on the experimental treatment in the delta-adjusted PMMs and be similar to that of control participants in the control-based PMMs. These PMMs generally yield more conservative estimates than the MAR-based analyses. While the MAR-based analysis estimates the efficacy (de jure estimand) of the treatment when the drug is taken as directed, the MNAR analysis assesses the effectiveness (de facto estimand) of the treatment as actually taken.[8] Please refer to Kenward[11] for a review on the concept of the controlled imputation and to Tang[3,12,13] for computational algorithms for continuous outcomes. The controlled imputation methods have been extended to time-to-event endpoints[14] and recurrent events.[4,15] An addendum to ICH E9 has been proposed relating to estimands and sensitivity analyses.[16]

Extension of the controlled imputation to longitudinal binary and ordinal outcomes is not straightforward partially because there is no natural multivariate distribution for such endpoints. In this paper, we propose 2 modeling approaches for implementing the controlled imputation for binary and ordinal data. One approach is built on a sequence of logistic regressions, the other approach uses the multivariate probit model. The proposed methods are potentially very useful as sensitivity analysis in a longitudinal clinical trial with the binary or ordinal outcome as the primary or key efficacy endpoints (the data are often collected at a number of prespecified visits).

Novel Markov chain Monte Carlo (MCMC) algorithms are designed for multiple imputations (MI) in both approaches. A critical feature of the controlled imputation is that the data prior to dropout are assumed to follow the same distribution as that in the MAR-based model (it implies that the intermittent missing data are MAR). Therefore, the marginal posterior distribution of the model parameters and the intermittent missing data will be the same under MAR and MNAR, and unrelated to the data after dropout. The basic idea underlying the MCMC is similar to that for continuous outcomes.[3,13] We can firstly design an algorithm to draw the model parameters and the intermittent missing data from their marginal posterior distribution, and this can be efficiently achieved via the monotone data augmentation (MDA) scheme. The missing data after dropout are imputed from their posterior distribution given the model parameters and the data prior to dropout after the MDA algorithm converges.

The MDA algorithm iterates between an imputation I-step, in which the intermittent missing data are imputed given the current draw of the model parameters, and a posterior P-step, which draws the model parameters from their posterior distribution given the current imputed monotone missing data. The MDA algorithm is a collapsed MCMC sampling scheme,[17,18] in which the missing data after dropout are integrated out of the posterior distribution, so the dimension of the parameter space (missing outcomes are treated as additional parameters) is greatly reduced.[3,12] As shown in Tang,[3] the MDA algorithm tends to converge faster with smaller autocorrelation between posterior samples than an FDA algorithm, in which both data prior to and after dropout are imputed in the I-step.

In the sequential regression, the model parameters are drawn from their posterior distribution using the Metropolis-Hastings (MH) sampler instead of a Gibber sampler, which is unlike the MDA algorithm for multivariate normal outcomes.[3] In the data augmentation (DA) algorithm[19] for the multivariate probit model, some constraints are placed on the covariance matrix of the latent variables or other parameters to ensure identifiability. It is not easy to sample a covariance matrix under these restrictions.[20,21] An MDA algorithm is constructed under alternative parameter constraints, and its convergence is accelerated via parameter expansion (PX).[22]

The rest of the paper is organized as follows. Section 2 describes the sequential regression–based controlled imputation procedure. In Section 3, we introduce the controlled imputation on basis of the multivariate probit model. Section 4 conducts simulation to assess the performance of these methods and compares them with the MI approach via fully conditional specification (FCS) and the single imputation methods such as the last observation carried forward (LOCF) and baseline observation carried forward (BOCF). The proposed methods are illustrated by the analysis of the National Institute of Mental Health (NIMH) schizophrenia trial in Section 5.

2 | SEQUENTIAL LOGISTIC REGRESSION APPROACH

We consider a 2-arm trial. Suppose n participants are randomly assigned to the experimental ($g_i = 1$) or control ($g_i = 0$) treatment. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})'$ be the outcomes at p postbaseline visits, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iQ})'$ the covariates for participant i. We set $x_{i1} \equiv 1$ for the intercept and $x_{iQ} = g_i$ for the treatment status. In general, $\boldsymbol{y}_i$'s are partially observed. Let s_i be the dropout pattern according to the time of the last observation. We have $s_i = 0$ for participants with no postbaseline assessment, and $s_i = p$ for completers. Let $\boldsymbol{\pi}_g = (\pi_{g0}, \ldots, \pi_{gp})'$ be the marginal pattern probabilities within each arm.

Let $\boldsymbol{y}_{io}, \boldsymbol{y}_{im_i}$, and $\boldsymbol{y}_{im_d}$ denote respectively the observed data, intermittent missing data, and missing data after dropout for participant i. Let $\vec{\boldsymbol{y}}_{it} = (y_{i1}, \ldots, y_{it})$. Then, $\boldsymbol{y}_{io}$ ($\boldsymbol{y}_{im_i}$) is the observed (missing) part of $\vec{\boldsymbol{y}}_{is_i}$. Let Y_o, Y_{m_i}, and Y_{m_d} denote respectively the observed outcomes, intermittent missing data, and the missing data after dropout from all participants, and $Y = (Y_o, Y_{m_i}, Y_{m_d})$. Without loss of generality, we sort the data so that participants in pattern s are arranged before participants in pattern t if $s > t$. Suppose y_{ij} is not missing in the first $n_j = n_{0j} + n_{1j}$ participants after filling in the intermittent missing data, where n_{gj} is the number of participants in pattern j, treatment group g.

2.1 | MDA algorithm

In the sequential regression approach, the joint distribution of $(y_{i1}, \ldots, y_{ip})$ can be factored as

$$f(y_{i1}, \ldots, y_{ip}) = \prod_{j=1}^{s_i} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j) \prod_{j=s_i+1}^{p} g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j), \tag{1}$$

where $\boldsymbol{z}_{ij} = (\boldsymbol{x}_i', \vec{\boldsymbol{y}}_{i,j-1}')'$, $\boldsymbol{z}_{i1} = \boldsymbol{x}_i$, and $f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$ and $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ are respectively the distributions of the observed and missing outcomes conditioning on the history. The observed data are modeled using the logistic regression or proportional odds model. We will discuss how to specify the missing data distribution $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ in Section 2.2. There might be additional parameters $\tilde{\boldsymbol{\theta}}_j$'s in $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ to capture deviation from MAR. It generally assumes that $\tilde{\boldsymbol{\theta}}_j$'s are fixed or follow a known distribution since they cannot be inferred from the observed data.

The likelihood for the complete data is given by

$$\mathcal{L}_c(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p|Y_o, Y_{m_i}, Y_{m_d}) = \left[\prod_{g=0}^{1}\prod_{j=0}^{p} \pi_{gj}^{n_{gj}}\right] \mathcal{L}_o(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p|Y_o, Y_{m_i}) \left[\prod_{j=1}^{p}\prod_{i=n_j+1}^{n} g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)\right].$$

where $\mathcal{L}_o(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p|Y_o, Y_{m_i}) \propto \prod_{j=1}^{p}\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$ is the likelihood for the augmented monotone data (Y_o, Y_{m_i}).

We assume independent priors for $\boldsymbol{\theta}_j$'s and $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1)$. That is, $\psi(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, \boldsymbol{\pi}_0, \boldsymbol{\pi}_1) = \psi(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1)\prod_{j=1}^{p}\psi(\boldsymbol{\theta}_j)$. Throughout, we let $\psi(\cdot)$ and $\Psi(\cdot|\cdot)$ denote respectively the prior and posterior distributions. The joint posterior distribution of $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, \boldsymbol{\pi}_0, \boldsymbol{\pi}_1, Y_{m_i}, Y_{m_d})$ can be written as

$$\Psi(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, \boldsymbol{\pi}_0, \boldsymbol{\pi}_1, Y_{m_i}, Y_{m_d}|Y_o) \propto \Psi(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1|Y_o)\Psi(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, Y_{m_i}|Y_o) \left[\prod_{j=1}^{p}\prod_{i=n_j+1}^{n} g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)\right], \tag{2}$$

where $\Psi(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1|Y_o) \propto \psi(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1)\prod_{g=0}^{1}\prod_{j=0}^{p}\pi_{gj}^{n_{gj}}$, and the marginal posterior distribution of $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, Y_{m_i})$ is

$$\Psi(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, Y_{m_i}|Y_o) \propto \prod_{j=1}^{p}\left[\pi(\boldsymbol{\theta}_j)\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)\right]. \tag{3}$$

An MDA algorithm (labeled as a) can be designed based on Equation 2.

1a. Draw $\boldsymbol{\theta}_j$'s and Y_{m_i}'s from their marginal posterior distribution defined in Equation 3.

   P. Draw $\boldsymbol{\theta}_j$ from $\Psi(\boldsymbol{\theta}_j|Y_o, Y_{m_i}) \propto \pi(\boldsymbol{\theta}_j)\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$ for $j = 1, \ldots, p$.
   I. Impute the intermittent missing data $\boldsymbol{y}_{im_i} \sim \prod_{j=1}^{s_i} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$ for participant $i = 1, \ldots, n$.

2a. Impute the missing data $\boldsymbol{y}_{im_d}$ after dropout sequentially from $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ for participant $i = 1, \ldots, n$.
3a. Draw $\boldsymbol{\pi}_g$'s from $\Psi(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1|Y_o)$.

Step 3a can be ignored for the purpose of missing data imputation since $\boldsymbol{\pi}_g$'s are independent of $\boldsymbol{\theta}_j$'s and Y_{m_i}'s in the posterior distribution. Step 2a does not affect step 1a and can be implemented after the posterior samples in step 1a reach

the stationary distribution. Step 1a defines a MDA scheme, in which only the intermittent missing data are imputed during the I-step. The MDA algorithm imputes fewer missing data in each iteration and tends to be more efficient than an FDA algorithm that imputes both $\boldsymbol{y}_{im_d}$'s and $\boldsymbol{y}_{im_i}$'s during the I-step and draws $\boldsymbol{\theta}_j$'s from its posterior distribution

$$\Psi(\boldsymbol{\theta}_j|Y_o, Y_{m_i}, Y_{m_d}) \propto \psi(\boldsymbol{\theta}_j)\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)\prod_{i=n_j+1}^{n} g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$$

given the (imputed) complete data for the reasons given in Section 1. Furthermore, the posterior distribution $\Psi(\boldsymbol{\theta}_j|Y_o, Y_{m_i}, Y_{m_d})$ in the FDA algorithm is more complex than that $\Psi(\boldsymbol{\theta}_j|Y_o, Y_{m_i})$ in the MDA algorithm.

In the sequential regression, we use the following proportional odds model for ordinal outcomes with K levels

$$\gamma_{ij_k} = \Pr(y_{ij} \le k|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j) = \text{expit}\left(c_{j_k} + \boldsymbol{\alpha}_j'\boldsymbol{x}_i + \sum_{t=1}^{j-1}\beta_{jt}y_{it}\right) \tag{4}$$

for $k = 1, \ldots, K-1$, where $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jQ})'$, $c_{j_1} < \ldots < c_{j_{K-1}}$, and $\text{expit}(a) = \exp(a)/[1+\exp(a)]$. If the proportional odds assumption does not hold, a different logistic regression[23] may be used. In model (4), we treat the outcomes at previous visits as continuous covariates, but they can also be coded as categorical covariates. We set $c_{j_1} = 0$ since it will be absorbed into the intercept. We reparameterize c_{j_k}'s as $d_{j_k} = \log(c_{j_k} - c_{j_{k-1}})$ [ie, $c_{j_k} = \sum_{t=2}^{k}\exp(d_{j_t})$] to ensure $c_{j_k} > c_{j_{k-1}}$. For the binary endpoint, y_{ij}'s are coded as 1 or 2, and model (4) reduces to

$$\Pr(y_{ij} = 1|y_{i1}, \ldots, y_{ij-1}) = \text{expit}\left(\boldsymbol{\alpha}_j'\boldsymbol{x}_i + \sum_{t=1}^{j-1}\beta_{jt}y_{it}\right).$$

Let $\boldsymbol{\theta}_j = (d_{j_2}, \ldots, d_{j_{K-1}}, \boldsymbol{\alpha}_j', \beta_{j1}, \ldots, \beta_{j,j-1})'$ and $\tau_{ij_k} = \Pr(y_{ij} = k|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$. Let $d^*_{j_k} = \exp(d_{j_k})$ for $k \le j$ and 0 if $k > j$. The score statistic and the Fisher information matrix for model (4) are given respectively by

$$\mathbf{U}(\boldsymbol{\theta}_j) = \sum_{i=1}^{n_j}\frac{\partial\tau_{ij_{y_{ij}}}}{\partial\boldsymbol{\theta}_j} \text{ and } \mathbf{I}(\boldsymbol{\theta}_j) = \sum_{i=1}^{n_j}\sum_{k=1}^{K}\tau_{ij_k}^{-1}\left[\frac{\partial\tau_{ij_k}}{\partial\boldsymbol{\theta}_j}\right]^{\otimes 2}, \tag{5}$$

where $\partial\gamma_{ij_k}/\partial\boldsymbol{\theta}_j = \gamma_{ij_k}(1-\gamma_{ij_k})[d^*_{j_2}, \ldots, d^*_{j_{K-1}}, \boldsymbol{x}_i', y_{i1}, \ldots, y_{ik-1}]'$, and

$$\frac{\partial\tau_{ij_k}}{\partial\boldsymbol{\theta}_j} = \begin{cases} \frac{\partial\gamma_{ij_1}}{\partial\boldsymbol{\theta}_j} & \text{at } k = 1, \\ \frac{\partial\gamma_{ij_k}}{\partial\boldsymbol{\theta}_j} - \frac{\partial\gamma_{ij_{k-1}}}{\partial\boldsymbol{\theta}_j} & \text{at } k = 2, \ldots, K-1, \\ -\frac{\partial\gamma_{ij_{K-1}}}{\partial\boldsymbol{\theta}_j} & \text{at } k = K. \end{cases}$$

We set the prior as $\boldsymbol{\theta}_j \sim N(v_j, R_j)$. When $R_j^{-1} \to \boldsymbol{0}$, the prior becomes flat, ie, $\psi(\boldsymbol{\theta}_j) \propto 1$. We use the MH sampler to draw $\boldsymbol{\theta}_j$'s from their posterior distribution given Y_o and the current imputed Y_{m_i} in the P-step: (1) Sample a candidate $\boldsymbol{\theta}_j^* \sim N[\boldsymbol{\theta}_j + \mu_j(\boldsymbol{\theta}_j), \Sigma_j(\boldsymbol{\theta}_j)]$ given the current point $\boldsymbol{\theta}_j$; (2) accept the move $\boldsymbol{\theta}_j \to \boldsymbol{\theta}_j^*$ with probability A_j, and otherwise, keep $\boldsymbol{\theta}_j$ unchanged, where $\Sigma_j(\boldsymbol{\theta}_j) = [\mathbf{I}(\boldsymbol{\theta}_j) + R_j^{-1}]^{-1}$, $\mu_j(\boldsymbol{\theta}_j) = \Sigma_j(\boldsymbol{\theta}_j)[\mathbf{U}(\boldsymbol{\theta}_j) + R_j^{-1}v_j]$, $\phi(x|\mu, \Sigma)$ is the probability density function for $x \sim N(\mu, \Sigma)$ and

$$A_j = \min\left\{1, \frac{\phi[\boldsymbol{\theta}_j|\mu_j(\boldsymbol{\theta}_j^*), \Sigma_j(\boldsymbol{\theta}_j^*)]\,\pi(\boldsymbol{\theta}_j^*)\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j^*)}{\phi[\boldsymbol{\theta}_j^*|\mu_j(\boldsymbol{\theta}_j), \Sigma_j(\boldsymbol{\theta}_j)]\,\pi(\boldsymbol{\theta}_j)\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)}\right\}.$$

The MH sampler generalizes the algorithm of Gamerman[24] for the generalized linear model by allowing nonlinear predictors in the model, and it does not require any parameter tuning. In our numerical examples, the acceptance rate is typically >65% due to the similarity between the proposal density $N[\mu_j(\boldsymbol{\theta}_j), \Sigma_j(\boldsymbol{\theta}_j)]$ and the posterior density of $\boldsymbol{\theta}_j$.

Alternatively, the Gibbs sampler may be used to sample $\boldsymbol{\theta}_j$'s. In large samples, one can find the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_j$ using participants in pattern j and draw $\boldsymbol{\theta}_j \sim N(\hat{\boldsymbol{\theta}}_j, [\mathbf{I}(\hat{\boldsymbol{\theta}}_j)]^{-1})$ since the posterior distribution of $\boldsymbol{\theta}_j$'s can be closely approximated by the asymptotic normal distribution of the MLE.[25,26] In smaller samples, the large

sample approximation is generally not adequate, and it takes a little extra effort to obtain samples from the exact posterior distribution by using the rejection sampling or sampling-importance-resampling technique.[25-27] We recommend the MH sampler since it can be computationally expensive to find the MLEs at each MCMC iteration.

Below, we describe how to impute $\boldsymbol{y}_{im_i}$'s given $\boldsymbol{y}_{io}$'s and the current draw of $\boldsymbol{\theta}_j$'s in the I-step. Suppose n_{im} observations are missing prior to dropout, and h_i is the index of the first missing observation for participant i. There are $K^{n_{im}}$ possible combinations of $\boldsymbol{y}_{im_i}$ (denoted by $\boldsymbol{y}^l_{im_i}$, $l = 1, \ldots, K^{n_{im}}$). Set $\boldsymbol{y}_{im_i} = \boldsymbol{y}^l_{im_i}$ with probability $\alpha_l / \sum_{l=1}^{K^{n_{im}}} \alpha_l$, where $\alpha_l = \prod_{j=h_i}^{s_i} f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \boldsymbol{y}_{im_i} = \boldsymbol{y}^l_{im_i})$.

2.2 | Imputations of missing data after dropout under MAR and MNAR

As mentioned in Section 2.1, the missing data after dropout can be imputed sequentially from $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ given the draw of the model parameters $\boldsymbol{\theta}_j$'s and imputed intermittent missing data $\boldsymbol{y}_{im_i}$'s after the MDA algorithm converges. Under MAR, $g(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j, \tilde{\boldsymbol{\theta}}_j)$ is identical to $f(y_{ij}|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j)$.

Below, we describe 2 MNAR mechanisms. In the delta-adjusted PMMs, the missing data distribution is

$$\Pr(y_{ij} \le k|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j) = \text{expit}\left[c_{j_k} + \sum_{q=1}^{\text{Q}-1}\left(\alpha_{jq} + \Delta^{\alpha}_{sj_q}\right)x_{iq} + (\delta_j + \Delta_{sj})g_i + \sum_{t=1}^{j-1}\left(\beta_{jt} + \Delta^{\beta}_{sj_t}\right)y_{it}\right] \tag{6}$$

at $j > s$ for participants in pattern s, where $\delta_j = \alpha_{j\text{Q}}$. The specification of model (6) is similar to that for continuous outcomes.[10] To reduce the number of sensitivity parameters, we set $\Delta^{\alpha}_{sj_q}$'s and $\Delta^{\beta}_{sj_t}$'s to 0 and place the restriction that $\Delta_{s,s+1} = \ldots = \Delta_{sp} = \Delta$, or that $\Delta_{s,s+1} = \Delta$ and $\Delta_{sj} = 0$ at $j \ge s+2$. Model (6) reduces to

$$\Pr(y_{ij} \le k|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j) = \text{expit}\left[c_{j_k} + \sum_{q=1}^{\text{Q}-1}\alpha_{jq}x_{iq} + (\delta_j + \Delta_{s_ij})g_i + \sum_{t=1}^{j-1}\beta_{jt}y_{it}\right]. \tag{7}$$

Model (7) implies MAR in the control group, and that in the experimental arm, the log odds of being in better health status after dropout are reduced compared to those who remain in the trial ($\Delta < 0$ if lower scores on y_{ij}'s indicate better health. Otherwise, $\Delta > 0$). In model (7), it is impossible to estimate Δ from the observed data. The tipping point analysis[3,11] is often used. We assume Δ is known and repeat the MI inference at a sequence of increasing Δ values to find the tipping point Δ, at which the treatment effect becomes insignificant. The MAR-based analysis is said to be robust if the tipping point is large and deemed clinically implausible. The tipping point can be found analytically for continuous outcomes[28] but may not exist for binary or ordinal outcomes. An example is given in Section 4. Section 6 will discuss that the tipping point analysis can also be performed by applying the delta adjustment in both treatment groups.[29]

The copy reference (CR) procedure[7,8] is a control-based PMM that formulates the missing data distribution among dropouts in the experimental arm on basis of the statistical behavior of control participants

$$\Pr(y_{ij} \le k|\boldsymbol{z}_{ij}, \boldsymbol{\theta}_j) = \text{expit}\left[c_{j_k} + \sum_{q=1}^{\text{Q}-1}\alpha_{jq}x_{iq} + \sum_{t=1}^{j-1}\beta_{jt}y_{it}\right] \text{ for } j > s_i. \tag{8}$$

In the MI analysis, one uses a standard method to analyze each imputed dataset and then combines the results from the M datasets for inference via Rubin's rule.[30] Let $\hat{\boldsymbol{\varpi}}^{(h)}$ denote the estimate of the parameters of interest and $\hat{W}^{(h)}$ the variance for dataset h. The MI estimate of $\boldsymbol{\varpi}$ and the associated Rubin's variance are given respectively by $\bar{\boldsymbol{\varpi}} = M^{-1}\sum_{m=1}^{M}\hat{\boldsymbol{\varpi}}^{(m)}$ and $\text{var}_R(\bar{\boldsymbol{\varpi}}) = \bar{W} + (1 + H^{-1})\bar{B}$, where $\bar{W} = M^{-1}\sum_{m=1}^{M}\hat{W}^{(m)}$ is the within-imputation variance, $\bar{B} = (M-1)^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\varpi}}^{(m)} - \bar{\boldsymbol{\varpi}})^{\otimes 2}$ is the between-imputation variance, and $a^{\otimes 2} = aa'$.

3 | MULTIVARIATE PROBIT MODEL

The univariate and multivariate probit models, popularized by Albert and Chib[31] and Chib and Greenberg,[19] provide an alternative approach for modeling binary and ordinal outcomes. It assumes that there exists latent variables w_{ij}'s and ordered thresholds $-\infty = c_0 < c_1 < \ldots < c_K = \infty$ such that $y_{ij} = k$ if $c_{k-1} < w_{ij} \le c_k$. The methods in Sections 2 and 3 are unrelated, so the parameters in the 2 sections may have different interpretations.

We assume that the latent variables follow a multivariate normal distribution

$$\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})' \sim N([\boldsymbol{\alpha}_1'\boldsymbol{x}_i, \ldots, \boldsymbol{\alpha}_p'\boldsymbol{x}_i]', \Sigma). \tag{9}$$

Let $\Sigma = L\Lambda L'$, where $\Lambda = \text{diag}(\gamma_1^{-1}, \ldots, \gamma_p^{-1})$, $U = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ -\beta_{21} & 1 & \ldots & 0 \\ -\beta_{p1} & \ldots & -\beta_{p,p-1} & 1 \end{bmatrix}$ and $L = U^{-1}$. We can rearrange Equation 9 as[3,28]

$$w_{ij} = \sum_{q=1}^{Q} \underline{\alpha}_{jq} x_{iq} + \sum_{t=1}^{j-1} \beta_{jt} w_{it} + \varepsilon_{ij} = \boldsymbol{\theta}_j' \boldsymbol{z}_{ij} + \varepsilon_{ij} \text{ for } j = 1, \ldots, p. \tag{10}$$

where $\boldsymbol{z}_{ij} = (\boldsymbol{x}_1', w_{i1}, \ldots, w_{ij-1})'$, $\underline{\boldsymbol{\alpha}}_j = (\underline{\alpha}_{j1}, \ldots, \underline{\alpha}_{jQ})' = \boldsymbol{\alpha}_j - \sum_{t=1}^{j-1} \beta_{jt}\boldsymbol{\alpha}_t$, $\boldsymbol{\theta}_j = (\underline{\boldsymbol{\alpha}}_1', \beta_{j1}, \ldots, \beta_{jj-1})'$, $\varepsilon_{ij} \sim N(0, \gamma_j^{-1})$, $w_{ij} \in B_{ij}$, $B_{ij} = (c_{k-1}, c_k]$ if $y_{ij} = k$, and $B_{ij} = (-\infty, \infty)$ for missing y_{ij}.

We set $c_1 = 0$ since it will be absorbed into the intercept α_1 in model (10). For binary outcomes, it is common to restrict the covariance matrix Σ to be a correlation matrix to ensure parameter identifiability,[19] and it can be challenging to design MCMC algorithms under such restrictions.[20,21] We impose alternative constraints by setting $\gamma_1 = \ldots = \gamma_p = 1$ for binary endpoints, and $\gamma_1 = 1$ for ordinal outcomes.

3.1 | The MDA algorithm

We construct MDA algorithms for the multivariate probit model and use the PX technique[22] to accelerate the convergence of the algorithm. In the MDA algorithm, only the latent variables prior to the dropout are imputed during the I-step, and the latent variables after dropout are imputed after the MDA algorithm converges. The missing binary or ordinal responses are imputed by comparing the imputed latent variables with the posterior samples of the thresholds $(c_1, c_2 \ldots, c_{K-1})$ after the convergence of the MDA algorithm.

In addition to the benefits given in Section 1, another advantage of the MDA algorithm is that the LDL decomposition of the covariance matrix is naturally provided in the algorithm,[3] which makes it computationally simple to impute the latent variables after the dropout visit.

3.1.1 | MCMC for binary outcomes

An MDA algorithm draws posterior samples of $(w_{i1}, \ldots, w_{is_i})$'s and $\boldsymbol{\theta}_j$'s under the constraint $\gamma_1 = \ldots = \gamma_p = 1$. We use the PX scheme[22] to speed up the convergence of the MDA algorithm at the cost of a minor increase in the per iteration computational burden. We treat γ_j's as the expansion parameters and introduce new latent variables $\tilde{w}_{ij} = w_{ij}/\sqrt{\gamma_j}$. The PX-MDA algorithm draws posterior samples of $(\tilde{w}_{ij}$'s, $\boldsymbol{\theta}_j$'s, γ_j's) or (w_{ij}'s, $\boldsymbol{\theta}_j$'s, γ_j's). For the univariate probit model,[22,32] the PX scheme has been shown to be more efficient than the routine DA algorithm and the reparametrization of Nandram and Chen.[33]

Suppose the prior for γ_j is $\mathcal{GA}(a_j, b_j)$ with density $\pi(\gamma_j) \propto \gamma_j^{a_j - 1} \exp(-b_j\gamma_j)$, and the prior for $\boldsymbol{\theta}_j$ is $N(v_j, R_j)$. When $R_j^{-1} \to \boldsymbol{0}$, the prior for $\boldsymbol{\theta}_j$ becomes flat. The posterior distributions are provided in Appendix A. The PX-MDA algorithm (labeled as b) can be described as follows:

1b. Impute $\tilde{\boldsymbol{w}}_{io} = (\tilde{w}_{i1}, \ldots, \tilde{w}_{is_i})|\boldsymbol{\theta}_j$'s, γ_j's, Y_o. This can be achieved by drawing $\gamma_{j_0} \sim \mathcal{GA}(a_j, b_j)$, imputing $\boldsymbol{w}_{io} = (w_{i1}, \ldots, w_{is_i})|\boldsymbol{\theta}_j$'s, Y_o at $\gamma_1 = \ldots = \gamma_p = 1$, and setting $\tilde{w}_{ij} = w_{ij}/\sqrt{\gamma_{j_0}}$.

2b. Sample $(\gamma_j, \boldsymbol{\theta}_j)|Y_o, \tilde{\boldsymbol{w}}_{io}$'s for $j = 1, \ldots, p$. This is equivalent to drawing $\gamma_j|Y_o, \tilde{\boldsymbol{w}}_{io}$'s, setting $w_{ij}^* = \sqrt{\gamma_j}\tilde{w}_{ij} = \sqrt{r_j}w_{ij}$, and sampling $\boldsymbol{\theta}_j|Y_o, \boldsymbol{w}_{io}^*$'s, where $r_j = \gamma_j/\gamma_{j_0}$ and $\boldsymbol{w}_{io}^* = (w_{i1}^*, \ldots, w_{is_i}^*)$.

Algorithm b can be equivalently written as the following algorithm b^*, but algorithm b is not feasible when $a_j \to 0$ since it is impossible to draw γ_{j_0} in step 1b.

1b*. Impute $\boldsymbol{w}_{io}|\boldsymbol{\theta}_j$'s, $\boldsymbol{y}_{io}$ at $\gamma_1 = \ldots = \gamma_p = 1$ for $i = 1, \ldots, n$. The conditional distribution of $\boldsymbol{w}_{io} = (w_{i1}, \ldots, w_{is_i})$ given $\boldsymbol{\theta}_j$'s and $\boldsymbol{y}_{io}$ is a multivariate normal distribution truncated to the region $B_{i1} \times \ldots \times B_{is_i}$. The Gibbs sampler of Li and Ghosh[34] is used to update $\varepsilon_{ij} = w_{ij} - \boldsymbol{\theta}_j'\boldsymbol{z}_{ij} (j = 1, \ldots, s_i)$ by conditioning on other ε_{ik}'s ($k \neq j, k \leq s_i$), and this method tends to be more efficient than sampling $w_{ij} (j \leq s_i)$ directly by conditioning on other w_{ik} ($k \neq j, k \leq s_i$) via the Gibbs sampler.

2b*. Draw r_j from Equation A2 in Appendix A. Calculate $\boldsymbol{w}_{io}^* = \sqrt{r_j}\boldsymbol{w}_{io}$.

3b*. Draw $\boldsymbol{\theta}_j|Y_o, \gamma_j$'s, $\tilde{\boldsymbol{w}}_{io}$'s or equivalently $\boldsymbol{\theta}_j|\boldsymbol{w}_{io}^*$'s from Equation A3 in Appendix A.

Algorithm b^* does not depend on the prior parameters b_j's. Setting $a_j \to 0$ is equivalent to putting a Haar measure prior[22] $\pi(g_j) = g_j^{-1}$ on $g_j = \sqrt{1/\gamma_j}$ or setting $\pi(\gamma_j) \propto \gamma_j^{-1}$. The movement in each PX-MDA iteration can be represented as $\boldsymbol{\theta}_{j_{\text{old}}}$'s $\to \boldsymbol{w}_{io}$'s $\to \boldsymbol{w}^*_{io}$'s $\to \boldsymbol{\theta}_{j_{\text{new}}}$'s, and the stationary distributions of $\boldsymbol{w}_{io}$ and $\boldsymbol{w}^*_{io}$ are the same. Ignoring step $2b^*$ by setting $\boldsymbol{w}^*_{io} = \boldsymbol{w}_{io}$ leads to a routine MDA algorithm. The addition of step $2b^*$ by adjusting $\boldsymbol{w}_{io}$'s to $\boldsymbol{w}^*_{io}$'s allows $\boldsymbol{\theta}_j$'s to move more freely and makes the Markov chain converge faster. All arguments are similar to that of Liu and Wu.[22]

3.1.2 | MCMC for ordinal outcomes

The parameter restriction is $\gamma_1 = 1$. In the PX scheme, we treat γ_1 as an expansion parameter and define $(\tilde{\boldsymbol{w}}_{io}\text{'s}, \tilde{c}_k\text{'s}) = (\boldsymbol{w}_{io}\text{'s}, c_k\text{'s})/\sqrt{\gamma_1}$. Suppose the priors are $\pi(\gamma_j) \propto \gamma_j^{-1}$, $\boldsymbol{\theta}_j \sim N(v_j, R_j)$ and $\pi(c_2, \ldots, c_{K-1}) \propto 1$. The posterior distributions are given in Appendix B.

Below is the PX-MDA algorithm (labeled as *c*). It can also be derived using the generalized Gibbs sampler of Liu and Sabatti.[32]

1*c*. Impute $\boldsymbol{w}_{io}|\boldsymbol{\theta}_j\text{'s}, \gamma_j\text{'s}, \boldsymbol{y}_{io}$ at $\gamma_1 = 1$ for $i = 1, \ldots, n$, and update $c_k \sim \text{uniform}(\max(\max_{y_{ij}=k-1} w_{ij}, c_{k-1}), \min(\min_{y_{ij}=k} w_{ij}, c_{k+1}))$ for $2 \le k \le K-1$.
2*c*. Draw r_1 from Equation B2 in Appendix B as $a_1 \to 0$. Compute $\boldsymbol{w}^*_{io} = \sqrt{r_1}\boldsymbol{w}_{io}$ and $c^*_k = \sqrt{r_1}c_k$.
3*c*. Draw $\gamma_j (j \ge 2)$ and $\boldsymbol{\theta}_j$'s from Equation B3 in Appendix B.

3.2 | Missing data imputations after dropout

This section discusses how to impute the latent variables w_{ij}'s after dropout under MAR and MNAR, which is similar to that for continuous endpoints.[3,13] The missing response both before ($\boldsymbol{y}_{im_i}$'s) and after ($\boldsymbol{y}_{im_d}$'s) dropout can be easily imputed by comparing w_{ij}'s with the thresholds $(c_1, c_2 \ldots, c_{K-1})$.

Under MAR, w_{ij} $(j > s_i)$ can be imputed sequentially from

$$w_{ij} = \sum_{q=1}^{\text{Q}} \underline{\alpha}_{jq} x_{iq} + \sum_{t=1}^{j-1} \beta_{jt} w_{it} + \varepsilon_{ij}, \tag{11}$$

or in matrix form as

$$(w_{is+1}, \ldots, w_{ip})' = L_{s+1,p}\left(\left[\underline{\mu}_{i,s+1_s} + \varepsilon_{i,s+1}, \ldots, \underline{\mu}_{ip_s} + \varepsilon_{ip}\right]'\right)$$

for participants in pattern $s < p$, where $\underline{\mu}_{ij_s} = \sum_{q=1}^{\text{Q}} \underline{\alpha}_{jq} x_{iq} + \sum_{t=1}^{s} \beta_{jt} w_{it}$, $\varepsilon_{ij} \sim N(0, \gamma_j^{-1})$ and $L_{s+1,p}$ is the lower right $(p-s)\times(p-s)$ submatrix of $L = U^{-1}$. Recall the restriction that $\gamma_1 = \ldots = \gamma_p = 1$ for binary endpoints, and $\gamma_1 = 1$ for ordinal outcomes.

Below, we describe 3 PMMs under MNAR. The delta-adjusted PMM assumes that in the experimental arm, the mean of w_{ij} after dropout becomes worse compared with participants who remain in the study[3,9,10]:

$$w_{ij} = \sum_{q=1}^{\text{Q}} \underline{\alpha}_{jq} x_{iq} + \sum_{t=1}^{j-1} \beta_{jt} w_{it} + \Delta_{s_i j} g_i + \varepsilon_{ij} \text{ for } j > s_i. \tag{12}$$

We set $(\Delta_{s,s+1}, \ldots, \Delta_{sp}) = (\Delta, \ldots, \Delta)$, or $(\Delta, 0, \ldots, 0)$.

In CR, participants in the experimental arm after dropout have the same future statistical behavior as control participants with identical historical outcomes[7,8]

$$w_{ij} = \sum_{q=1}^{\text{Q}-1} \underline{\alpha}_{jq} x_{iq} + \sum_{t=1}^{j-1} \beta_{jt} w_{it} + \varepsilon_{ij} \text{ for } j > s_i. \tag{13}$$

For participants in pattern $s < p$, $(w_{i,s+1}, \ldots, w_{ip})$'s in the delta-adjusted and CR imputation procedures can be obtained respectively by subtracting $-L_{s+1,p}[\Delta_{s,s+1}, \ldots, \Delta_{sp}]' g_i$ and $L_{s+1,p}[\delta_{s+1}, \ldots, \delta_p]' g_i$ from the imputed values under MAR,[13] where $\delta_j = \alpha_{j\text{Q}}$.

The jump to reference (J2R) is another popular control-based PMM.[8] It assumes that all treatment benefits are gone immediately after participants discontinue the experimental treatment. Let $\mu^{\text{c}}_{ij} = \sum_{q=1}^{\text{Q}-1} \alpha_{jq} x_{iq}$ and $[\delta_1, \ldots, \delta_p] = L[\underline{\delta}_1, \ldots, \underline{\delta}_p]'$. Under MAR, the distribution of $(w_{i1}, \ldots, w_{ip})'$ is $N([\mu^{\text{c}}_{i1} + \delta_1 g_i, \ldots, \mu^{\text{c}}_{ip} + \delta_p g_i]', \Sigma)$. In J2R,

$$(w_{i1}, \ldots, w_{ip})' \sim N([\mu^c_{i1} + \delta_1 g_i, \ldots, \mu^c_{is} + \delta_s g_i, \mu^c_{i,s+1}, \ldots, \mu^c_{ip}]', \Sigma)$$

for participants in pattern s, and $w_{ij}(j > s_i)$ can be imputed by subtracting δ_j from the imputed values under MAR.[13] It is difficult to specify J2R in the sequential regression since it needs the unconditional treatment effect at visit j that is unadjusted for $(y_{i1}, \ldots, y_{ij-1})$.

4 | SIMULATION

We conduct simulations to assess the performance of the proposed procedures, and compare them with the FCS, LOCF and BOCF methods. We consider two scenarios, and simulate $H = 5000$ datasets of size $n = 300$ (150 participants per arm) in each scenario. In scenario I, the data are generated from the multivariate probit model $w_{i0} \sim N(0, 1)$, $w_{i1}|w_{i0} \sim N(0.5w_{i0}+0.8g_i, 1)$, $w_{i2}|w_{i0}, w_{i1} \sim N(0.5+0.3w_{i0}-0.4w_{i1}+0.4g_i, 1)$, $w_{i3}|w_{i0}, w_{i1}, w_{i2} \sim N(0.3-0.4w_{i0}+0.5w_{i1}+0.3w_{i2}+0.6g_i, 1)$, $y_{ij} = 1$ if $w_{ij} \le 0$, and $y_{ij} = 2$ if $w_{ij} > 0$. Pattern is determined according to the probability $\Pr(s_i = 0) = \text{expit}(0.5y_{i0} - 4.8)$, $\Pr(s_i = 1|s_i \ge 1) = \text{expit}(0.5y_{i0} + 0.3y_{i1} - 4)$, and $\Pr(s_i = 2|s_i \ge 2) = \text{expit}(0.3y_{i0} + 0.5y_{i1} + 0.4y_{i2} - 3.6)$. The proportions of participants in patterns $0 - 3$ are roughly $(1.78\%, 6.10\%, 15.18\%, 76.94\%)$. Intermittent missing data are created by setting $y_{ij}(1 \le j < s_i)$ to be missing with a 3% chance.

In scenario II, the data are simulated sequentially from $\Pr(y_{i0} = 2) = 0.6$, $\Pr(y_{i1} = 2|y_{i0}) = \text{expit}(0.2 + 0.3y_{i0} + 0.5g_i)$ and $\Pr(y_{i2} = 2|y_{i0}, y_{i1}) = \text{expit}(0.2 + 0.2y_{i0} + 0.5y_{i1} + 0.8g_i)$. Pattern is generated according to $\Pr(s_i = 0) = \text{expit}(0.3y_{i0} - 2.5)$, and $\Pr(s_i = 1|s_i \ge 1) = \text{expit}(0.3y_{i0} + 0.5y_{i1} - 2.4)$. The proportions of participants in patterns $0 - 2$ are approximately $(11.81\%, 22.82\%$, and $65.38\%)$. We set $y_{ij}(1 \le j < s_i)$ to be missing with a 5% chance. In both scenarios, no baseline y_{i0} is missing, and $\boldsymbol{x}_i = (1, y_{i0}, g_i)'$.

The simulated data are imputed using both the sequential regression and multivariate probit model under MAR and MNAR. The imputation model differs from the true model when the sequential regression (multivariate probit model) is used to fit the data in scenario I (II). We use noninformative priors by setting $R_j \to \mathbf{0}$ and $a_j \to 0$ in all examples in Sections 4 and 5. For each simulated dataset, $M = 100$ posterior samples are collected every 20th iteration after a burn-in period of 3000 iterations in the sequential regression, and $M = 100$ datasets are imputed at every 50th iteration after a burn-in period of 5000 iterations in the multivariate probit model. Each imputed dataset is analyzed using the logistic regression by visit. The results are summarized in Table 1.

To enable assessment of the bias in the MI estimates, we present 2 full data estimates, in which the missing data after dropout are generated according to the assumed MAR or MNAR mechanisms. Column 1 of Table 1 presents the "pooled full data estimate" based on a huge dataset combining observations from all $H = 5000$ replications. Column 2 of Table 1 displays the "averaged full data estimate," which is the average of $H = 5000$ complete data estimates. The 2 full data estimates differ even if we increase the number of replications H but do not increase the sample size n within each dataset. Data in scenario I (II) are not analyzed with the delta-adjusted imputation using the sequential regression (multivariate probit model) because it is difficult to assess the bias of MI estimates in the 2 cases.

The sequential regression and multivariate probit model produce very similar MI estimates under MAR in both scenarios, and the estimates in CR are also quite close especially in scenario II. The MI results from the multivariate probit model in scenario I and that from the sequential regression in scenario II are generally close to the "averaged full data estimate." We also perform simulations under higher dropout rate while assuming the same observed data distribution as scenarios I and II. The bias in the parameter estimates slightly increases when about 40% to 50% participants discontinue the study early. The results are omitted because they follow similar patterns to that in Table 1.

In PMMs, the imputation and analysis models differ in that the data are imputed by assuming that the joint distribution of $\boldsymbol{y}_i$ varies by pattern under a relevant MNAR assumption but analyzed by a standard method that follows the intent-to-treat principle. The PMMs provide estimates of the effectiveness (de facto estimand) of the treatment as actually taken. As shown in Table 1, the proposed PMMs yield more conservative treatment effect estimates than the MAR-based imputation since discontinued participants in the experimental arm are assumed to have worse response profiles in these PMMs than under MAR. Furthermore, Rubin's rule overestimates the sampling variance of the treatment effect estimates in J2R and CR but not under the MAR and delta-adjusted imputations. For example, in scenario I, Rubin's variance averaged over the 5000 replications for the J2R treatment effect estimates is 0.115 while the sample variance of the J2R estimates is only 0.046. Rubin's variance estimate is generally biased in the presence of uncongeniality between the imputation and analysis models,[35] and the phenomenon is similar to that for continuous outcomes.[28]

TABLE 1 Comparison of parameter estimates at last visit based on the proposed methods by simulation

				Imputation Via Multivariate Probit Model					Imputation Via Sequential Logistic Regression				
				MI	Rubin's Variance			Sample	MI	Rubin's Variance			Sample
	Parameter	est1[a]	est2[b]	Estimate	Between	Within	Total	Variance[c]	Estimate	Between	Within	Total	Variance[c]
Scenario I: data generated from the multivariate probit model													
MAR imputation													
	Intercept	−1.098	−1.114	−1.115	0.054	0.240	0.295	0.308	−1.113	0.054	0.241	0.295	0.308
	Treatment	−1.626	−1.662	−1.677	0.032	0.119	0.151	0.161	−1.681	0.033	0.119	0.152	0.162
	y_{i0}	0.297	0.302	0.301	0.024	0.090	0.114	0.117	0.299	0.024	0.090	0.114	0.117
MNAR imputation													
CR	Intercept	−1.141	−1.157	−1.170	0.064	0.225	0.289	0.298	−1.190	0.068	0.221	0.289	0.300
	Treatment	−1.283	−1.306	−1.293	0.036	0.094	0.130	0.064	−1.193	0.032	0.089	0.122	0.052
	y_{i0}	0.326	0.331	0.337	0.028	0.083	0.112	0.113	0.350	0.030	0.081	0.112	0.115
J2R	Intercept	−1.175	−1.190	−1.195	0.067	0.216	0.284	0.290					
	Treatment	−1.067	−1.085	−1.090	0.030	0.085	0.115	0.046					
	y_{i0}	0.348	0.353	0.354	0.029	0.079	0.109	0.110					
Delta[d]	Intercept	−1.141	−1.158	−1.157	0.063	0.229	0.293	0.305					
	Treatment	−1.341	−1.367	−1.407	0.044	0.102	0.146	0.150					
	y_{i0}	0.326	0.331	0.329	0.027	0.085	0.113	0.115					
Scenario II: data generated from the sequential logistic regression													
MAR imputation													
	intercept	−0.969	−0.991	−1.007	0.140	0.324	0.465	0.480	−1.011	0.140	0.325	0.466	0.484
	treatment	−0.845	−0.866	−0.887	0.061	0.125	0.187	0.196	−0.889	0.060	0.126	0.187	0.197
	y_{i0}	−0.230	−0.229	−0.227	0.054	0.114	0.169	0.175	−0.228	0.054	0.115	0.169	0.177
MNAR imputation													
CR	Intercept	−1.021	−1.043	−1.056	0.154	0.300	0.456	0.459	−1.057	0.159	0.301	0.461	0.473
	Treatment	−0.513	−0.524	−0.532	0.044	0.103	0.148	0.054	−0.522	0.044	0.103	0.147	0.053
	y_{i0}	−0.194	−0.192	−0.195	0.061	0.104	0.165	0.169	−0.198	0.063	0.104	0.167	0.174
Delta[e]	Intercept	−1.000	−1.022						−1.043	0.155	0.310	0.466	0.486
	Treatment	−0.647	−0.662						−0.669	0.071	0.113	0.184	0.196
	y_{i0}	−0.207	−0.206						−0.207	0.060	0.108	0.168	0.176

Abbreviations: CR, copy reference; MAR, missing at random; MI, multiple imputations; MNAR, missing not at random.

[a] Estimate from a dataset with 1 500 000 observations pooled over all $H = 5000$ replications, where missing data after dropout are generated according to the true mechanism.

[b] Average of $H = 5000$ complete data estimate, where missing data after dropout are generated according to the true mechanism.

[c] Sample variance of $H = 5000$ MI estimates.

[d] An adjustment of −0.5 is applied to the mean of the latent variables at all visits after dropout.

[e] An adjustment of 0.5 is applied to the log odds at all visits after dropout.

TABLE 2 Comparison of parameter estimates at last visit using full conditional specification (FCS), last observation carried forward (LOCF), and baseline observation carried forward (BOCF) by simulation

	FCS Imputation				Single Imputation			
	MI	Rubin's Variance			LOCF		BOCF	
Parameter	Estimate	Between	Within	Total	Estimate	Variance	Estimate	Variance
Scenario I: data generated from the multivariate probit model								
Intercept	−1.110	0.055	0.240	0.295	−0.708	0.215	0.806	0.197
Treatment	−1.666	0.034	0.118	0.152	−1.271	0.093	−1.057	0.082
y_{i0}	0.299	0.024	0.090	0.115	0.019	0.082	−0.954	0.081
Scenario II: data generated from the sequential logistic regression								
Intercept	−1.004	0.143	0.321	0.466	0.130	0.246	1.958	0.244
Treatment	−0.876	0.063	0.124	0.188	−0.609	0.096	−0.478	0.096
y_{i0}	−0.223	0.055	0.113	0.169	−0.851	0.093	−2.042	0.108

Abbreviation: MI, multiple imputations.

We also analyze the simulated data using the FCS, LOCF, and BOCF methods. The results are reported in Table 2. The FCS procedure is a popular MI approach for multivariate data containing nonnormal or mixed-type outcomes.[36] It imputes data on a variable-by-variable basis by specifying a logistic regression model for each incomplete variable conditional on all other variables. The algorithm generally converges within 10 to 20 iterations.[36,37] For each simulated dataset, we use SAS Proc MI to impute $M = 100$ datasets after a burn-in period of 100 iterations. The mean FCS estimate over the 5000 replications is quite close to the results from the 2 proposed procedures under MAR.

In LOCF and BOCF, the missing values after dropout are replaced by the last observed value (the baseline is used to impute the missing value if there is no postbaseline assessment) and baseline value, respectively. As shown in Table 2, the variance estimate from the 2 single imputation methods are much lower than that in the MI approaches. Furthermore, the 2 methods tend to yield biased estimates due to unrealistic assumptions about the missing data.[5,6]

5 | ANALYSIS OF THE NIMH SCHIZOPHRENIA TRIAL

We illustrate the proposed methods by analyzing the NIMH schizophrenia collaborative study.[38] Participants were randomized into placebo or 1 of the 3 antipsychotic treatments. Participants from the 3 active treatments were pooled together[38] due to similar treatment effects. The dataset published at http://hedeker.people.uic.edu/SCHIZX1.DAT.txt contains 108 participants on placebo and 329 participants on the active treatment.

Item 79 (severity of illness) of the inpatient multidimensional psychiatric scale was collected at baseline and weeks 1, 3, and 6 and analyzed as an ordinal outcome with 4 levels (1 = normal or borderline mentally ill, 2 = mildly or moderately ill, 3 = markedly ill, and 4 = severely or extremely ill) or binary outcome (1 = normal to mildly ill and 2 = moderately to extremely ill) using the random-effects probit or ordinal regression models.[38,39] The proportions of participants in patterns 0 to 3 are $(0, 16.7, 18.5, 64.8)\%$ and $(0.9, 8.2, 10.3, 80.5)\%$ respectively in the placebo and active arms. In addition, 21 participants have intermittent missing data.

We assess the treatment effect under various missing data mechanisms. Baseline y_{i0} is not included as a covariate since it is highly correlated with the intercept (about 98.6% participants are moderately to extremely ill). We set $\boldsymbol{x}_i = (1, g_i)'$.

For the sequential regression, the trace plots and autocorrelation function plots indicate that the MDA algorithm a reaches approximate convergence within few hundreds of iterations. The autocorrelation decays quickly to 0. The lag-10 autocorrelation is below 0.1 for nearly all model parameters. In the multivariate probit models, it takes several thousands of steps for the PX-MDA algorithm to converge. The use of the PX scheme in algorithms b^* and c greatly increases the efficiency of the algorithms. Without such a scheme, the autocorrelations for a few parameters remain above 0.72 at lag 200 for both binary and ordinal outcomes, which reduce to the 0.2 to 0.3 range in the PX-MDA algorithms.

We impute $M = 1000$ datasets from every 100th iteration after a burn-in period of 10 000 iterations. Each imputed dataset is analyzed using the logistic regression by visit. The results are summarized in Table 3. Similar results are obtained when we try different initial values for the model parameters, vary the prior parameter values (ie, use weakly informative prior) or use longer burn-in periods. The multivariate probit model and sequential regression approach yield similar treatment effect estimates for both binary and ordinal endpoints under the MAR and CR imputations. The MNAR imputations

TABLE 3 Estimated treatment effects by visit and associated Rubin's variance for the National Institute of Mental Health (NIMH) schizophrenia trial using the proposed methods

		Imputation Via Multivariate Probit Model					Imputation Via Sequential Logistic Regression				
		MI	Rubin's Variance				MI	Rubin's Variance			
Imputation	Visit	Estimate	Between	Within	Total	*t*	Estimate	Between	Within	Total	*t*
Binary outcomes											
MAR imputation											
	1	0.632	0.005	0.121	0.127	1.775	0.636	0.006	0.122	0.128	1.780
	2	1.290	0.017	0.097	0.114	3.823	1.332	0.019	0.099	0.118	3.878
	3	1.440	0.025	0.061	0.085	4.933	1.418	0.025	0.060	0.085	4.868
MNAR imputation											
CR	1	0.627	0.005	0.122	0.127	1.761	0.631	0.006	0.122	0.128	1.768
	2	1.230	0.015	0.097	0.112	3.669	1.246	0.017	0.100	0.116	3.655
	3	1.284	0.020	0.060	0.080	4.540	1.231	0.019	0.060	0.079	4.381
J2R	1	0.627	0.005	0.122	0.127	1.761					
	2	1.202	0.015	0.097	0.113	3.581					
	3	1.172	0.019	0.060	0.079	4.172					
Delta[a]	1	0.620	0.005	0.122	0.127	1.743	0.628	0.006	0.122	0.127	1.759
	2	1.146	0.016	0.097	0.113	3.403	1.262	0.018	0.099	0.117	3.682
	3	1.108	0.028	0.060	0.088	3.744	1.260	0.024	0.060	0.084	4.349
Delta[b]	1	0.620	0.005	0.122	0.127	1.742	0.624	0.006	0.122	0.128	1.748
	2	1.123	0.016	0.097	0.113	3.340	1.163	0.018	0.100	0.118	3.389
	3	0.879	0.022	0.060	0.082	3.063	0.862	0.022	0.060	0.082	3.008
Ordinal outcomes											
MAR imputation											
	1	0.788	0.001	0.043	0.044	3.762	0.791	0.001	0.043	0.044	3.771
	2	1.331	0.005	0.044	0.050	5.967	1.330	0.005	0.044	0.050	5.958
	3	1.857	0.012	0.048	0.060	7.587	1.855	0.013	0.048	0.062	7.465
MNAR imputation											
CR	1	0.780	0.001	0.043	0.044	3.726	0.784	0.001	0.043	0.044	3.737
	2	1.249	0.005	0.044	0.049	5.644	1.241	0.005	0.044	0.049	5.627
	3	1.612	0.010	0.046	0.056	6.833	1.589	0.010	0.046	0.056	6.717
J2R	1	0.780	0.001	0.043	0.044	3.726					
	2	1.200	0.005	0.044	0.049	5.442					
	3	1.454	0.009	0.045	0.054	6.250					
Delta[a]	1	0.766	0.001	0.043	0.044	3.661	0.782	0.001	0.043	0.044	3.727
	2	1.044	0.005	0.043	0.048	4.783	1.250	0.005	0.044	0.049	5.622
	3	1.081	0.010	0.043	0.053	4.708	1.667	0.013	0.047	0.060	6.818
Delta[b]	1	0.764	0.001	0.043	0.044	3.651	0.766	0.001	0.043	0.044	3.658
	2	0.971	0.004	0.043	0.046	4.510	0.971	0.004	0.043	0.046	4.505
	3	0.852	0.006	0.042	0.048	3.871	0.863	0.007	0.042	0.049	3.882

Abbreviations: CR, copy reference; J2R, jump to reference; MAR, missing at random; MI, multiple imputations; MNAR, missing not at random.

[a] An adjustment of −1 is applied to the log odds at all visits after dropout in the sequential regression, and an adjustment of 2 is applied to the mean of the latent variables at all visits after dropout in the multivariate probit model.

[b] The response is set to "4" for ordinal outcome and "2" for binary outcome among active participants after dropout.

produce more conservative estimates than the MAR imputations, and the J2R is more conservative than CR. As discussed in Section 2, the delta-adjusted imputations are often implemented via the tipping point strategy. For this example, the tipping point does not exist. The tests of the treatment effect at visits 2 and 3 are still significant when we set all missing responses due to dropout in active participants to the worst values. These analyses indicate that participants treated with the antipsychotic treatments showed significantly more improvement in the IMPS 79 severity scores than participants on placebo and that the conclusion is robust to deviations from the MAR mechanism.

Table 4 displays the results from the FCS and LOCF imputations. Since y_{i0} is not included in the analysis, the BOCF imputation is not performed. The FCS estimates are close to the MAR results (reported in Table 3) based on the

TABLE 4 Estimated treatment effects by visit and associated variance for the National Institute of Mental Health (NIMH) schizophrenia trial using full conditional specification (FCS) and last observation carried forward (LOCF) methods

	FCS Imputation[a]					Single Imputation		
	MI	Rubin's Variance				LOCF		
Visit	Estimate	Between	Within	Total	t	Estimate	Variance	Z [b]
Binary outcomes								
1	0.633	0.006	0.122	0.128	1.770	0.718	0.132	1.978
2	1.313	0.020	0.098	0.119	3.811	1.296	0.101	4.070
3	1.381	0.024	0.059	0.084	4.778	1.579	0.068	6.066
Ordinal outcomes								
1	0.795	0.001	0.043	0.044	3.829	0.828	0.045	3.922
2	1.318	0.006	0.044	0.050	5.886	1.353	0.045	6.387
3	1.883	0.016	0.049	0.065	7.378	1.906	0.469	8.669

Abbreviation: MI, multiple imputations.
[a] $M = 100$ datasets are imputed after a burn-in period of 100 iterations using SAS Proc MI.
[b] The Z statistic is calculated as the square root of the Wald chi-square statistic

multivariate probit model and sequential regression. The LOCF approach is not necessarily conservative. It gives slightly larger treatment effect estimate than FCS at visit 3 for both binary and ordinal outcomes and seems to underestimate the variance of the treatment effect. We provide SAS programs for the analysis in the Supporting Information.

6 | DISCUSSION

We describe the controlled imputation for longitudinal binary and ordinal outcomes via the sequential logistic regression and multivariate probit models. Efficient MCMC algorithms are developed for the 2 approaches using the MDA and/or PX techniques. We focus on several commonly used imputation strategies such as the CR, J2R, and delta-adjusted imputations. The proposed methods show good performance in our limited simulation.

There are many potential ways to formulate the missing data distribution. One may assume that the future statistical behaviors of outcomes after dropout vary according to both the dropout pattern and reason. In the delta-adjusted PMM, the response in participants in the experimental arm become worse after dropout, but the data in the control arm are MAR. Alternatively, we may apply the delta adjustment in both arms (Δ_E for the experimental treatment, Δ_c for the control treatment) and perform the MI inference over a wide range of values for (Δ_E, Δ_c) to find the region in which the treatment comparison becomes insignificant.[29] Heat maps are useful for visualizing the result.[2] The imputation method under these potential new MNAR mechanisms will be similar to that described in Sections 2.2 and 3.2 if they assume the same observed data distribution as that under MAR.

In the sequential regression, we use the logistic regression to model the outcome at each visit conditioning on the historical outcomes. As suggested by 1 referee, an alternative approach is to use the sequential (univariate) probit models since it is easier to develop Gibbs samplers in the univariate probit model (the idea is similar to that in Section 3). We expect similar results from the 2 sequential regression approaches because the probit and logit functions approximately satisfy $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-\frac{x^2}{2})dx \approx \frac{1}{1+\exp(-1.6a)}$ over a wide range of a; see Pingel[40] and references therein. We choose the logistic regression since it is more commonly used in medical research due to its easy interpretation.

It is interesting to see that under MAR, the sequential regression, multivariate probit models, and FCS give similar results in both the simulation study and the analysis of the NIMH schizophrenia trial. The FCS procedure generally performs well under MAR,[36,37,41] and the fact is demonstrated for longitudinal ordinal outcomes by simulation in Donneau et al.[42] Although MNAR imputation can also be implemented in FCS (eg, multiply or shift the imputed values by a constant amount[41]), it is usually unknown to which multivariate distribution the algorithm converges, or the algorithm may not converge.[36,37,41] It would be challenging to interpret the FCS result under MNAR even for multivariate normal outcomes.[13] Compared to FCS, the proposed procedures allow some explicit and clinically plausible MNAR assumptions about the disease evolution after dropout that can be easily understood by the clinicians.

It is possible to extend the sequential regression approach to longitudinal outcomes of mixed types. With this extension, one may incorporate different types of intermediate outcomes into the imputation model to improve the accuracy of the imputed values or impute several clinical endpoints simultaneously.[43] Some automated MCMC algorithms shall be designed to handle these complex situations. In our numerical examples, we focus on large clinical trials and use the

noninformative prior, which has little influence on the inference. When the sample size is small or when the number of variables p is relatively large compared to the sample size n, the imputation model can be overparameterized, and the parameter estimates may become unstable. One possible solution is to use (weakly) informative priors or Bayesian penalized methods.[44,45] Alternatively, constraints may be placed on the parameters or their priors to reduce the number of free parameters. For example, in the multivariate probit model, the dimension of covariance parameters can be reduced by using a structured covariance matrix or by using prior distribution families for the unstructure covariance matrix that allow the data to determine a compromise between unstructured and parametric matrices.[43,46] Further research shall be conducted on these topics.

ACKNOWLEDGEMENT

We would like to thank the associate editor and 2 referees for their helpful suggestions that improve the quality of the work.

ORCID

Yongqiang Tang http://orcid.org/0000-0002-8997-8421

REFERENCES

1. ICH E9. Statistical principles for clinical trials: ICH harmonized tripartite guideline. *Stat Med*. 1999;18:1905-1942.
2. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. *Stat Med*. 2014;33:4170-4185.
3. Tang Y. An efficient monotone data augmentation algorithm for multiple imputation in a class of pattern mixture models. *J Biopharm Stat*. 2017;27:620-638.
4. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharm Stat*. 2014;13:258-264.
5. CHMP. *EMA Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99)*. London: CHAMP; 2010.
6. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press; 2010.
7. Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*. 1996;52:1324-1333.
8. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat*. 2013;23:1352-1371.
9. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, FL: Chapman and Hall; 2008.
10. Wang C, Daniels M. A note on MAR, identifying restrictions, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*. 2011;67:810-818.
11. Kenward MG. Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clin Invest*. 2015;5:311-320.
12. Tang Y. An efficient monotone data augmentation algorithm for Bayesian analysis of incomplete longitudinal data. *Stat Probab Lett*. 2015;104:146-152.
13. Tang Y. An efficient multiple imputation algorithm for control-based and delta-adjusted pattern mixture models using SAS. *Stat Biopharm Res*. 2017;9:116-125.
14. Lu K, Li D, Koch GG. Comparison between two controlled multiple imputation methods for sensitivity analyses of time-to-event data with possibly informative censoring. *Stat Biopharm Res*. 2015;7:199-213.
15. Tang Y. Algorithms for imputing partially observed recurrent events with applications to multiple imputation in pattern mixture models. *J Biopharm Stat*. 2017. https://doi.org/10.1080/10543406.2017.1333999.
16. ICH E9 Steering Committee. E9(R1): addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials; 2014.
17. Liu JS, Wong WH, Kong A. Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*. 1994;81:27-40.
18. Liu JS. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J Am Stat Assoc*. 1994;89:958-966.
19. Chib S, Greenberg E. Bayesian analysis of multivariate probit models. *Biometrika*. 1998;85:347-361.
20. Zhang X, Boscardin W, Belin. Sampling correlation matrices in Bayesian models with correlated latent variables. *J Comput Graphical Stat*. 2006;15:880-896.
21. Liu X, Daniels MJ. A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *J Comput Graphical Stat*. 2006;15:897-914.
22. Liu JS, Wu YN. Parameter expansion for data augmentation. *J Am Stat Assoc*. 1999;94:1264-1274.
23. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol*. 1997;26:1323-1333.

24. Gamerman D. Efficient sampling from the posterior distribution in generalized linear mixed models. *Stat Comput*. 1997;7:57-68.
25. Zeger SL, Karim MR. Generalized linear models with random effects; a Gibbs sampling approach. *J Am Stat Assoc*. 1991;86:79-86.
26. Raghunathan TE, Lepkowski JM, van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Method*. 2001;27:85-95.
27. Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling-resampling perspective. *The Am Stat*. 1992;46:84-88.
28. Tang Y. On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models. *Biometrics*. 2017. https://doi.org/10.1111/biom.12702.
29. Permutt T. Sensitivity analysis for missing data in regulatory submission. *Stat Med*. 2016;35:876-879.
30. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley Sons, Inc; 1987.
31. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc*. 1993;88:669-679.
32. Liu JS, Sabatti C. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*. 2000;87:353-369.
33. Nandram B, Chen M. Accelerating Gibbs sampler convergence in generalized linear models via a reparametrization. *J Stat Comput Simul*. 1996;45:129-144.
34. Li Y, Ghosh SK. Efficient sampling methods for truncated multivariate normal and student-t distribution subject to linear inequality constraints. *J Stat Theory Pract*. 2015;9:712-732.
35. Meng X. Multiple-imputation inference with uncongenial sources of input. *Stat Sci*. 1994;9:538-573.
36. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16:219-242.
37. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30:377-399.
38. Hedeker DR, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics*. 1994;50:933-944.
39. Gibbons RD, Hedeker D. Application of random-effects probit regression models. *J Consulting Clin Psychology*. 1994;62:285-296.
40. Pingel R. Some approximations of the logistic distribution with application to the covariance matrix of logistic regression. *Stat Probab Lett*. 2014;85:63-68.
41. van Buuren S, Groothuis-Oudshoon K. Mice: multivariate imputation by chained equations in R. *J Stat Software*. 2011;45:1-67.
42. Donneau AF, Mauer M, Lambert P, Molenberghs G, Albert A. Simulation-based study comparing multiple imputation methods for non-monotone missing ordinal data in longitudinal settings. *J Biopharm Stat*. 2015;25:570-601.
43. He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. *Stat Med*. 2014;33:2251-2262.
44. Park T, Casella G. The Bayesian LASSO. *J Am Stat Assoc*. 2008;103:681-686.
45. Mallick H, Yi N. Bayesian methods for high dimensional linear models. *J Biometrics Biostat*. 2013;S1:005. https://doi.org/10.4172/2155-6180.S1-005
46. Boscardin WJ, Weiss R. Models for the covariance matrix of multivariate longitudinal and repeated measures data. In: Proceedings of American Statistical Association 2001, Section on Bayesian Statistical Science: Alexandria, VA; 2001.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Tang Y. Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Statistics in Medicine*. 2018;37:1467–1481. https://doi.org/10.1002/sim.7583

APPENDIX A: POSTERIOR DISTRIBUTIONS FOR MULTIVARIATE PROBIT MODELS BINARY OUTCOMES

The conditional posterior distribution of $(\gamma_j, \boldsymbol{\theta}_j)$'s given Y_o and $\tilde{\boldsymbol{w}}_{io}$'s is

$$\prod_{j=1}^{p}\left\{\gamma_j^{a_j-1}\exp(-b_j\gamma_j)\gamma_j^{\frac{n_j}{2}}\exp\left[\frac{-\gamma_j\sum_{j=1}^{n_j}\left(\tilde{w}_{ij}-\underline{\boldsymbol{\alpha}}_j'\boldsymbol{x}_i\frac{1}{\sqrt{\gamma_j}}-\sum_{t=1}^{j-1}\tilde{w}_{it}\beta_{jt}\sqrt{\frac{\gamma_t}{\gamma_j}}\right)^2}{2}\right]\pi(\boldsymbol{\theta}_j)\right\}, \tag{A1}$$

$$\propto\prod_{j=1}^{p}\left\{\gamma_j^{a_j+\frac{n_j}{2}-1}\exp\left[-\gamma_j\left(\frac{\hat{S}_j}{\gamma_{j_0}}+b_j\right)\right]\exp\left[\frac{-(\boldsymbol{\theta}_j-\hat{\boldsymbol{\theta}}_j^*)'(Z_j^{*\prime}Z_j^*)(\boldsymbol{\theta}_j-\hat{\boldsymbol{\theta}}_j^*)}{2}\right]\pi(\boldsymbol{\theta}_j)\right\},$$

where $W_j = (w_{1j}, \dots, w_{n_j j})'$, $W_j^* = (w_{1j}^*, \dots, w_{n_j j}^*)' = \sqrt{r_j} W_j$, $\mathbf{1}_t$ is a $t \times 1$ vector of ones, $A_j = \text{diag}(\mathbf{1}_Q', \sqrt{r_1}, \dots, \sqrt{r_{j-1}})$, Z_j, and $Z_j^* = Z_j A_j$ are $n_j \times (Q + j - 1)$ matrices whose ith row are given respectively by $\boldsymbol{z}_{ij}'$ and $(\boldsymbol{x}_i', w_{i1}^*, \dots, w_{ij-1}^*)$, $\hat{\boldsymbol{\theta}}_j^* = (Z_j^{*\prime} Z_j^*)^{-1} Z_j^{*\prime} W_j^*$, and $\hat{S}_j$ is the sum of squared residuals from the regression of w_{ij} on $\boldsymbol{z}_{ij}$.

The marginal posterior distributions of γ_j's and r_j's can be derived from Equation A1:

$$\gamma_j | Y_o, \tilde{\boldsymbol{w}}_{io}\text{'s} \sim \mathcal{GA}\left(a_j + \frac{n_j}{2}, b_j + \frac{\hat{S}_j}{2\gamma_{j_0}}\right), r_j | Y_o, \tilde{\boldsymbol{w}}_{io}\text{'s} \sim \frac{\mathcal{GA}(a_j + n_j/2, 1)}{\mathcal{GA}(a_j, 1) + \hat{S}_j/2}, \text{ and } r_j | Y_o, \tilde{\boldsymbol{w}}_{io}\text{'s} \sim \frac{\chi^2_{n_j}}{\hat{S}_j} \text{ as } a_j \to 0. \tag{A2}$$

The posterior distribution of $\boldsymbol{\theta}_j$ is given by

$$\boldsymbol{\theta}_j | Y_o, \gamma_j, \tilde{\boldsymbol{w}}_{io}\text{'s} \sim N[\boldsymbol{\theta}_j^{**}, \Sigma_j^{**}], \tag{A3}$$

where $\Sigma_j^{**} = (A_j Z_j' Z_j A_j + R_j^{-1})^{-1}$ and $\boldsymbol{\theta}_j^{**} = \Sigma_j^{**}(\sqrt{r_j} A_j Z_j' W_j + R_j^{-1} v_j)$. If the prior for $\boldsymbol{\theta}_j$ is flat ($R_j^{-1} \to \mathbf{0}$), then $\boldsymbol{\theta}_j^{**} = \sqrt{r_j} A_j^{-1} B_1^{-1} B_2$, $\Sigma_j^{**} = A_j^{-1}(Z_j' Z_j)^{-1} A_j^{-1}$, and we can generate $\boldsymbol{\theta}_j$ as $\boldsymbol{\theta}_j = A_j^{-1} B_1^{-1}[\sqrt{r_j} B_2 + e_j]$, where e_j is a $(Q + j - 1) \times 1$ vector of standard normal random variables, $B = \begin{bmatrix} B_1 & B_2 \\ \mathbf{0} & c \end{bmatrix}$ is the Cholesky factor of $\begin{bmatrix} Z_j' Z_j & Z_j' W_j \\ W_j' Z_j & W_j' W_j \end{bmatrix} = B'B$, and $\hat{S}_j = c^2$.

APPENDIX B: POSTERIOR DISTRIBUTIONS FOR MULTIVARIATE PROBIT MODELS—ORDINAL OUTCOMES

The posterior distribution of ($\boldsymbol{\theta}_j$'s, γ_j's) given Y_o, $\tilde{\boldsymbol{w}}_{io}$'s and $\tilde{c}_k$'s can be derived by assuming γ_1 has the prior $\mathcal{GA}(a_1, b_1)$ and then taking the limit as $a_1 \to 0$

$$\begin{aligned}
&\gamma_1^{a_1 - 1} \exp(-b_1 \gamma_1) \prod_{t=2}^{p} \gamma_t^{-1} \prod_{t=1}^{p} \pi(\boldsymbol{\theta}_t) \prod_{i=1}^{n} f(\boldsymbol{w}_{io} | c_k\text{'s}, \boldsymbol{\theta}_j\text{'s}, \gamma_j\text{'s}) \left| \frac{\partial(\boldsymbol{w}_{io}\text{'s}, c_2, \dots, c_{K-1})}{\partial(\tilde{\boldsymbol{w}}_{io}\text{'s}, \tilde{c}_2, \dots, \tilde{c}_{K-2})} \right| \\
&\propto \gamma_1^{\frac{n_1 + K - 2}{2} + a_1 - 1} \exp\left[-\gamma_1 \left(\frac{\gamma_1 \sum_{j=1}^{n_1} \left(\tilde{w}_{i1} - \underline{\boldsymbol{\alpha}}_1' \boldsymbol{x}_i \frac{1}{\sqrt{\gamma_1}} \right)^2}{2} + b_1 \right) \right] \pi(\boldsymbol{\theta}_1) \\
&\prod_{t=2}^{p} \left\{ \gamma_1^{\frac{n_t}{2}} \gamma_t^{\frac{n_t - 1}{2}} \exp\left[-\frac{\gamma_1 \gamma_t \sum_{j=1}^{n_t} \left(\tilde{w}_{it} - \underline{\boldsymbol{\alpha}}_t' \boldsymbol{x}_i \frac{1}{\sqrt{\gamma_1}} - \sum_{s=1}^{t-1} \tilde{w}_{is} \beta_{ts} \right)^2}{2} \right] \pi(\boldsymbol{\theta}_t) \right\},
\end{aligned} \tag{B1}$$

where $\left| \frac{\partial(\cdot)}{\partial(\cdot)} \right|$ is the Jacobian transformation. As in algorithm b, we draw $\gamma_{1_0} \sim \mathcal{GA}(a_1, b_1)$, set $(\tilde{\boldsymbol{w}}_{io}, \tilde{c}_k) = (\boldsymbol{w}_{io}, c_k)/\sqrt{\gamma_{1_0}}$, and then derive the posterior distribution of $r_1 = \gamma_1/\gamma_{1_0}$

$$r_1 | \boldsymbol{w}_{io}\text{'s}, c_k\text{'s} \sim \frac{\mathcal{GA}(a_1 + \frac{n_1 + K - 2}{2}, 1)}{\mathcal{GA}(a_1, 1) + \frac{\hat{S}_1}{2}}, \text{ and } r_1 | \boldsymbol{w}_{io}\text{'s}, c_k\text{'s} \sim \frac{\chi^2_{n_1 + K - 2}}{\hat{S}_1} \text{ as } a_1 \to 0, \tag{B2}$$

where $\hat{S}_j$ is the SSR from the regression of w_{ij} on $\boldsymbol{z}_{ij}$. The posterior distributions of $\gamma_j (j \geq 2)$ and $\boldsymbol{\theta}_j$'s are given by

$$\gamma_j | \boldsymbol{w}_{io}^*\text{'s} \sim \mathcal{GA}\left(\frac{n_j}{2}, \frac{\hat{S}_j r_1}{2} \right), \text{ and } \boldsymbol{\theta}_j | \boldsymbol{w}_{io}^*\text{'s}, \gamma_j\text{'s} \sim N(\boldsymbol{\theta}_j^{**}, \Sigma_j^{**}), \tag{B3}$$

where $\boldsymbol{w}_{io}^* = \sqrt{r_1} \boldsymbol{w}_{io}$, $A_j = diag(\mathbf{1}_Q, \sqrt{r_1} \mathbf{1}_{j-1})$, $\Sigma_1^{**} = (X_1' X_1 + R_1^{-1})^{-1}$, $\boldsymbol{\theta}_1^{**} = \Sigma_1^{**}(\sqrt{r_1} X_1' W_1 + R_1^{-1} v_1)$, $\Sigma_j^{**} = (\gamma_j A_j Z_j' Z_j A_j + R_j^{-1})^{-1}$, and $\boldsymbol{\theta}_j^{**} = \Sigma_j^{**}(\sqrt{r_1} A_j Z_j' W_j + R_j^{-1} v_j)$ at $j \geq 2$.