Research Article

Statistics in Medicine

Received 28 July 2011, Accepted 10 February 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5359

A transition model for quality-of-life data with non-ignorable non-monotone missing data

Kaijun Liao,[a*†] Derek R. Freres[b] and Andrea B. Troxel[a]

In this paper, we consider a full likelihood method to analyze continuous longitudinal responses with non-ignorable non-monotone missing data. We consider a transition probability model for the missingness mechanism. A first-order Markov dependence structure is assumed for both the missingness mechanism and observed data. This process fits the natural data structure in the longitudinal framework. Our main interest is in estimating the parameters of the marginal model and evaluating the missing-at-random assumption in the Effects of Public Information Study, a cancer-related study recently conducted at the University of Pennsylvania. We also present a simulation study to assess the performance of the model.

Keywords: full likelihood; transition probability; selection model

1. Introduction

In a longitudinal study, subjects are observed as time progresses. A common problem is that repeated measurements are not fully observed because of missing responses or loss to follow-up. Individuals can move in and out of the observed data set, giving rise to a large class of distinct ‘non-monotone’ missingness patterns. The appropriate statistical methods differ according to the data structure and missing mechanism. When the missingness is missing completely at random (MCAR) or missing at random (MAR), data analysis is the most straightforward. Little and Rubin [1] and Allison [2] provide helpful terminology to describe missing-data mechanisms and a comprehensive overview of potential methods. Most approaches can be categorized as selection models, pattern-mixture models, or shared-parameter models depending on the factorization of the joint likelihood of the outcomes and missingness indicators. This article will focus on selection models.

Under the MCAR mechanism, the observed data can be viewed as a random subset of the complete data. With MAR data, the missingness mechanism depends only on observed quantities. Both mechanisms are termed ‘ignorable’ if the parameters in the two parts of the model are distinct. Unbiased parameter estimates can be guaranteed using generalized estimating equations, defined by Liang and Zeger [3] when the missingness mechanism is MCAR, and using weighted estimating equations defined by Robins and Rotnitzky [4] when the missingness mechanism is MAR. Neither method provides consistent unbiased estimators under informative dropout or non-ignorable (NI) missingess. The approaches to modeling longitudinal NI missing data depend on the data structure and type, variance–covariance structure, and proportion of missing data. Many proposed methods assume a multivariate Gaussian distribution for the outcomes, with different specifications of the covariance structure; these include Verbyla and Cullis [5], Richard and Lynn [6], Munoz *et al*. [7] and Diggle and Kenward [8].

Diggle and Kenward [8] proposed a likelihood-based method for continuous longitudinal outcomes with informative dropout, using a multivariate Gaussian distribution for the data and a logistic model for

[a] *Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
[b] *Annenberg School For Communication, University of Pennsylvania, Philadelphia, PA, USA*
**Correspondence to: Kaijun Liao, Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Drive, 1134 Blockley Hall, Philadelphia, PA 19104, USA.*
† *E-mail: kliao@mail.med.upenn.edu*

the probability of dropout. They allowed the missingness probability to depend on previous and current measurements and integrated the likelihood over the range of the unobserved values using numerical integration and iterative computations. Troxel *et al.* [9] extended the method to allow a non-monotone NI missingness mechanism. Their logistic model similarly allowed the probability of non-response to depend on the value of the current and/or previous measurement and assumed multivariate Gaussian underlying outcomes. They assumed a first-order Markov dependence structure to facilitate estimation.

Pseudolikelihood methods can also be used with NI non-monotone missingness to ease the computational burden of full likelihood approaches by setting a nuisance parameter at zero or some convenient value for estimation. Troxel *et al.* [10], Sinha *et al.* [11], and Parzen *et al.* [12] applied pseudolikelihood methods to binary data. Shen and Weissfeld [13] used a copula function to link the outcome and the missing-data mechanism. Troxel *et al.* [14] used an optimal weighted combination of two pseudolikelihoods to increase efficiency. Shardell and Miller [15] extended weighted estimating equations to allow both random and non-random missing-data mechanisms. Pattern-mixture models can also be used to model differences in outcome distributions over patterns of missing data; Birmingham and Fitzmaurice [16], Pauler *et al.* [17], and Shardell *et al.* [18] all proposed useful approaches. Tsonaka *et al.* [19] considered a semi-parametric shared-parameter model without assuming any parametric form for the random effects distribution. Robins [20] and Robins and Gill[21] proposed a semi-parametric method for non-monotone missing data where selection depends on unobserved variables.

Our method is an extension of the work of Troxel *et al.* [9] and is useful in situations with intermittent missing data from subjects moving in and out of the observed population; this happens commonly in longitudinal medical studies. Although the approach can be computationally intense with large numbers of assessments, it is very flexible with respect to the timing of assessments. We adopt a multivariate Gaussian distribution for the underlying data and a first-order Markov dependence structure. Instead of using logistic regression to model the missing mechanism, we propose a beta-binomial distribution for the probability of non-response; this helps to stabilize the estimation of the missingness probabilities and reduces multimodality in the likelihood, a common problem with many of the methods described above. The multivariate Polya distribution is a high-dimensional version of the beta-binomial distribution; the beta and binomial distributions correspond to Dirichlet and multinomial distributions, respectively, in the multivariate situation. Using this property, our approach can be easily extended to additional state of missingness, such as intermediate missingness, dropout, or even death if there is non-response due to death. Because of the gamma function and/or beta functions involved, closed-form maximum likelihood estimates are impractical. We propose the use of Gauss–Hermite quadrature as suggested in Liu and Pierce [22] to approximate the maximum likelihood and use the Broyden–Fletcher–Goldfarb–Shanno [23] algorithm to search for optimal solutions. The beta-binomial model provides better model fitting than a logistic model, especially for unbalanced sparse binary data.

We apply our approach to data from the Penn Center of Excellence in Cancer Communication Research. Effective communication between patients and physicians is very important in cancer treatment and throughout the healthcare system. Effective exchange of information between patients, physicians, healthcare systems, and the surrounding environment determines how active participants are within the healthcare system. Many studies show a link between isolation and worse outcomes [24], including shorter survival time, worse quality of life, and lower rates of participation in recommended treatment programs. Patient adherence to treatment is normally higher in those who actively seek information about their treatment and quality of life from multiple channels [25]. It is crucial to understand the relationship between patients, physicians, and the healthcare system, as well as the role of shared decision-making skills; how patients get, give, and discuss information and make healthcare decisions is important in cancer research, especially given the high demands that the healthcare system is facing.

The Effects of Public Information Study enrolled a total of 2010 patients diagnosed in 2005 with breast, colorectal, or prostate cancer selected from the Pennsylvania Cancer Registry. Subjects responded to at least one of three surveys, including 1520 patients who responded at wave 1, 1243 patients who responded at wave 2, and 1079 patients who responded at wave 3; these three surveys occurred at yearly intervals beginning in fall 2006. The American Association for Public Opinion Research response rates [26] for the primary sample were 68%, 64%, and 61% for the respective cancer groups [27]; intermitent missingness patterns were common. Surveys were mailed to all participants using Dillman's design method [28]. All patients were first mailed an introductory letter explaining the purpose of the study and including instructions; the surveys were mailed in a subsequent packet with a small monetary incentive ($3 or $5 for the short or long version of the survey). Reminder letters were sent after 2 weeks for subjects who did not return the survey. At follow-up assessments, contact was attempted with all patients, regardless of response to the prior year's survey. Patient consent was provided prior to participation, and the University of Pennsylvania Institutional Review Board reviewed and approved this study.

One of the study's research goals was to examine how the Patient–Clinician Information Engagement (PCIE) score affects cancer patients' attitudes and behaviors; in particular, researchers were interested in the amount of exercise the patients engaged in. For example, patients decide whether to increase exercise, to get radiation therapy, or to choose surgery after seeking and considering treatment information with their physicians; the decision-making process may be influenced by both medical and non-medical information. The PCIE score was designed to measure these constructs using eight items assessing whether, during the first few months of their cancer diagnosis, they had sought cancer treatment or quality-of-life information from their own or other physicians or from other sources. Each of the eight 'Yes/No' questions was transformed to a Z-score, and the average of the eight Z-scores formed the PCIE scale. We use the extent of exercise ('During an average week, how many days do you exercise?') as the primary outcome. The outcomes range from 0 to 7 by design; we treat these as continuous responses in this small interval. The Pearson correlation coefficients for the between-wave exercise scores ranged from 0.58 to 0.64. We use an unstructured correlation matrix in the data analysis and simulation sections, and we extend the correlation to AR(1), Exchangeable, and Toep(1) later in the simulation section for further model assessment.

We describe the proposed methods in Section 2 and illustrate these with an analysis of the PCIE data in Section 3. We present a simulation study to address the performance of the methods in Section 4. Section 5 provides a discussion and ideas for future work.

2. Methods and notation

2.1. Notation and underlying assumptions

Given a longitudinal data set, let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})'$ represent the vector of repeated measurements for subject i ($i = 1, \ldots, n$) with T measurement times. Let $\mathbf{X}_i$ be a vector of p covariates observed on the ith subject. Because the repeated measurements are not fully observed at each time point $t = (1, \ldots, T)$, define a vector of missingness indicators $\mathbf{R}_i = (R_{i1}, R_{i2}, \ldots, R_{iT})$ to correspond with the outcome vector $\mathbf{Y}_i = (\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}})$. Each element of R_i is defined as

$$R_{it} = \begin{cases} 0 & \text{if} \quad \text{missing} \\ 1 & \text{if} \quad \text{observed} \end{cases}.$$

For each subject, the full data are given by the repeated measurements and missingness indicators with joint distribution $L(\theta, \beta | \mathbf{Y}_i, \mathbf{R}_i, \mathbf{X}_i) \propto P(\mathbf{Y}_i, \mathbf{R}_i | \mathbf{X}_i, \theta, \beta)$. By partitioning $\mathbf{Y}_i$ into $(\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}})$, we can rewrite the joint likelihood in several ways.

A selection model would specify the joint distribution using the marginal distribution of the repeated outcomes and the conditional distribution of missing indicators:

$$P(\mathbf{Y}_i, \mathbf{R}_i | \mathbf{X}_i, \theta, \beta) = P(\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}} | \mathbf{X}_i, \beta) P(\mathbf{R}_i | \mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}}, \mathbf{X}_i, \theta).$$

A pattern-mixture model assumes that the full data have different distributions across strata determined by the pattern of missingness:

$$P(\mathbf{Y}_i, \mathbf{R}_i | \mathbf{X}_i, \theta, \beta) = P(\mathbf{R}_i | \mathbf{X}_i, \theta) P(\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}} | \mathbf{R}_i, \mathbf{X}_i, \beta).$$

A shared-parameter model assumes independence between the complete data and missing indicators conditional on a group of shared parameters $\boldsymbol{\gamma}$:

$$P(\mathbf{Y}_i, \mathbf{R}_i | \mathbf{X}_i, \theta, \beta) = \int P(\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{mis}} | \gamma_i, \mathbf{X}_i, \beta) P(\mathbf{R}_i | \gamma_i, \mathbf{X}_i, \theta) p(\gamma_i) \mathrm{d}\gamma_i.$$

In our study, we focus on selection models, which are a natural way to factor the joint likelihood function. The diagram below indicates the relationships among the variables graphically. Each arrow indicates the dependence among the nodes.

$$\begin{array}{ccccccccc} Y_{i1} & \longrightarrow & Y_{i2} & \cdots & Y_{i,T-1} & \longrightarrow & Y_{iT} \\ \downarrow & \cdots & \downarrow & \cdots & \downarrow & \cdots & \downarrow \\ R_{i1} & \longrightarrow & R_{i2} & \cdots & R_{i,T-1} & \longrightarrow & R_{iT} \end{array}$$

We adopt a model similar to Troxel *et al.* [9] and assume $\mathbf{Y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where the mean structure $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \cdots \mu_{iT})$ depends on a p-dimensional covariate vector $\mathbf{X}_i$. We also assume a first-order Markov dependence structure for both the full outcome data and the missingness indicators, so that $f(Y_{it}|Y_{i1}, Y_{i2}, \ldots Y_{i,t-1}) = f(Y_{it}|Y_{i,t-1})$ and $f(R_{it}|R_{i1}, R_{i2}, \ldots R_{i,t-1}) = f(R_{it}|R_{i,t-1})$. Let $\sigma_t^2 = \text{var}(Y_{it})$ and $\rho_t = \text{corr}(Y_{it}, Y_{i,t+1})$. Then we can denote the conditional likelihood as

$$Y_{it}|Y_{i,t-1} \sim N\left\{\mu_{it} + \rho_{t-1}\frac{\sigma_t}{\sigma_{t-1}}(Y_{i,t-1} - \mu_{i,t-1}), \sigma_t^2(1-\rho_{t-1}^2)\right\}.$$

For $T = 3$, the first-order ante-dependence structure is denoted as

$$\Sigma = \left\{\begin{array}{ccc} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 \end{array}\right\}.$$

2.2. *Missingness mechanism model*

Unlike other approaches to modeling the missingness mechanism, we are interested in the transition probability of the missingness indicators R_{it}. Conditional on each time t, the missingness mechanism becomes a two-state Markov chain. We model the transition probabilities $\pi_{jk} = \Pr(R_{it} = j|R_{i,t-1} = k, Y_{it}, X_{it})$, $j = 0, 1; k = 0, 1$ as

$$\begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

which satisfy the equation $\pi_{00} + \pi_{01} = \pi_{10} + \pi_{11} = 1$. We assume that the initial state is independent and define n_{ijk} as the number of times in the whole sequence that k is followed by j:

$$\begin{aligned} n_{ijk} &= \textstyle\sum_{t=1}^{T} I(R_{it} = j|R_{i,t-1} = k) \\ n_{ij.} &= \textstyle\sum_k n_{j,k}, \quad n_{i.k} = \sum_j n_{j,k}. \end{aligned}$$

Then the missingness mechanism can be written as

$$\begin{aligned} \mathbb{L}_i &= \pi_{00}^{n_{i00}}\pi_{01}^{n_{i01}}\pi_{10}^{n_{i10}}\pi_{11}^{n_{i11}} \\ &= \prod_{t=2}^{T}\prod_{j=0}^{1}\prod_{k=0}^{1}\pi_{jk}(t)^{I(R_{it}=j|R_{i,t-1}=k)}. \end{aligned}$$

This becomes a product of binomial distributions. To avoid the unstable estimation problems for binary outcomes near the boundary of the parameter space, we estimate the probability of missingness at each time t using a joint beta-binomial distribution instead of traditional estimation using logistic regression. From a Bayesian perspective, the beta distribution is the conjugate prior distribution for the parameters of the binomial distribution. The parameters of the beta distribution can be viewed as pseudocounts of 'response' and 'non-response' to be added to the observed counts of responses and non-responses.

Given time $t-1$, the missingness mechanism follows $(R_{it}|R_{i,t-1} = k) \sim \text{Bernoulli}(\pi_{ikt})$; we impose a beta distribution on the missingness probability, $\pi_{ikt} \sim \text{Beta}(a_{ikt}, b_{ikt})$. Then, we have

$$\begin{aligned} f(R_{it}|R_{i,t-1}, y_{it}, \pi) &= \prod_{k=0}^{1}\pi_{k1}^{I(R_{it}=1)I(R_{i,t-1}=k)}(1-\pi_{k1})^{[1-I(R_{it}=1)]I(R_{i,t-1}=k)} \\ f(\pi_{k1}|a_{k1}, b_{k1}) &= \frac{\Gamma(a_{k1}+b_{k1})}{\Gamma(a_{k1})\Gamma(b_{k1})} \times \pi_{k1}^{a_{k1}-1}(1-\pi_{k1})^{b_{k1}-1}. \end{aligned}$$

Integrating the π out, the mixture function can be expressed as

$$\begin{aligned} & f(R_{it}|R_{i,t-1}, a_{ik1}, b_{ik1}, y_{it}) \\ &= \int_0^1 f(R_{it}|R_{i,t-1} = k, \pi_{ikt}) f(\pi_{ikt}|a_{ikt}, b_{ikt}, y_{it}) \mathrm{d}\pi_{ikt} \\ &= \prod_{k=0}^{1}\frac{\Gamma(a_{ik1}+b_{ik1})}{\Gamma(a_{ik1})\Gamma(b_{ik1})} \times \frac{\Gamma(a_{ik1}+I(R_{it}=1)I(R_{i,t-1}=k))\Gamma(b_{ik1}+[1-I(R_{it}=1)]I(R_{i,t-1}=k))}{\Gamma(a_{ik1}+b_{ik1}+I(R_{i,t-1}=k))} \end{aligned}$$

with $a_{ik1} = \exp(\zeta_1 \mathbf{X_{it}} + \vartheta_1 \mathbf{Y_{it}} + \psi_1 \mathbf{R_{i,t-1}})$ and $b_{ik1} = \exp(\zeta_2 \mathbf{X_{it}} + \vartheta_2 \mathbf{Y_{it}} + \psi_2 \mathbf{R_{i,t-1}})$. However, the link function chosen could be different, resulting in a different missingness mechanism model. For given $R_{i,t-1} = 0$, the transition probability can be denoted as

$$P(R_{it} = l | R_{i,t-1} = 0, Y_{it}, X_{it}) : \; = \begin{cases} \frac{1}{1+\exp((\zeta_1-\zeta_2)\mathbf{X_{it}}+(\vartheta_1-\vartheta_2)\mathbf{Y_{it}})} & \text{if} \quad l = 1 \\ \frac{1}{1+\exp(-(\zeta_1-\zeta_2)\mathbf{X_{it}}-(\vartheta_1-\vartheta_2)\mathbf{Y_{it}})} & \text{if} \quad l = 0 \end{cases} .$$

For given $R_{i,t-1} = 1$,

$$P(R_{it} = l | R_{i,t-1} = 1, Y_{it}, X_{it}) : \; = \begin{cases} \frac{1}{1+\exp((\zeta_1-\zeta_2)\mathbf{X_{it}}+(\vartheta_1-\vartheta_2)\mathbf{Y_{it}}+(\psi_1-\psi_2))} & \text{if} \quad l = 1 \\ \frac{1}{1+\exp(-(\zeta_1-\zeta_2)\mathbf{X_{it}}-(\vartheta_1-\vartheta_2)\mathbf{Y_{it}}-(\psi_1-\psi_2))} & \text{if} \quad l = 0 \end{cases} .$$

Notice that if $\vartheta_1 - \vartheta_2 \neq 0$, then the missingness mechanism is indeed non-ignorable as the probability of missingness depends on the unobserved outcome Y_{it}. In practice, only the difference of the parameters is identifiable, not the individual parameters. We let $\zeta_c = \zeta_1 - \zeta_2$, $\vartheta_c = \vartheta_1 - \vartheta_2$ and $\psi_c = \psi_1 - \psi_2$ be the final parameters in the missingness mechanism model, where ζ_c is the coefficient of the covariates, ϑ_c is the coefficient of the current observed outcome y_{it}, and ψ_c is the coefficient of the previous missingness indicator $r_{i,t-1}$. Note that the covariates $\mathbf{X_{it}}$ can include effects for time and/or interaction terms between time and other variables, making it highly flexible and able to accommodate a wide range of effects on the missing-data probabilities.

The link function for parameters a_{ikt} and b_{ikt} of the beta distribution could be chosen as something other than a simple exponential function, and this will result in a different missingness mechanism model. The missingness mechanism model could be expanded similarly to a Dirichlet multinomial distribution from the current beta-binomial distribution when modeling more than two missingness states, such as 'observed', 'intermediate missing', and 'drop out'.

2.3. *Parameter estimation*

The observed joint likelihood function can be denoted as

$$\begin{aligned} \mathcal{L}_i(\mu, \Sigma, \beta) &= f(\mathbf{Y}_{i,\text{obs}}, \mathbf{R}_i) \\ &= \int \ldots \int f(\mathbf{Y}_{i,\text{obs}}, \mathbf{Y}_{i,\text{miss}}, \mathbf{R}_i) \mathrm{d}\mathbf{Y}_{i,\text{miss}} \\ &= \int \ldots \int f(Y_{i1}) f(R_{i1}|Y_{i1}) \prod_{t=2}^{T} f(Y_{it}|Y_{i,t-1}) f(R_{it}|R_{i,t-1}, Y_{it}) \mathrm{d}\mathbf{Y}_{i,\text{miss}}. \end{aligned}$$

There is no closed form for the observed likelihood function due to the complicated joint likelihood; a numerical integration method will be applied to approximate the likelihood function. The Gauss–Hermite quadrature rule is defined as

$$\begin{aligned} \int_{\mathbb{R}} f(t) \mathrm{d}\lambda(t) &= \int_{\mathbb{R}} f(t) w(t) \mathrm{d}t \\ &= \int_{\mathbb{R}} f(t) \exp(-t^2) \mathrm{d}t = \sum_{k=1}^{m} w_k f(\tau_k) + R_m(f) \end{aligned}$$

where m is the number of nodes, $\mathrm{d}\lambda(t) = w(t)\mathrm{d}t = \exp(-t^2)\mathrm{d}t$ is the measure with bounded or unbounded support on $\mathbb{R}$, w_k is the weight of the Gauss–Hermite quadrature rule, τ_k are the nodes (zero roots of the mth order Hermite polynomials), and $R_m(f)$ is the error term. The τ_k are symmetric about zero. The error term $R_m(f)$ will be zero if $f(t)$ is polynomial with degree less than $2m - 1$. Let $\phi(t; \mu, \sigma)$ be a normal density with mean μ and standard deviation σ. Then for any given function $f(t)$, we can approximate an integral as a summation following Liu and Pierce's [22] transformation

$$\int_{\infty}^{\infty} f(t)\phi(t; \mu, \sigma) \mathrm{d}t \simeq \sum_{i=1}^{m} \frac{w_i}{\sqrt{\pi}} f(\mu + \sqrt{2}\sigma\tau_k).$$

For $T = 3$, we list all possible data patterns and the joint likelihood function in the Appendix.

Our model is likelihood based, so maximum likelihood theory holds for parameter estimation. Letting $\eta = (\mu_1, \mu_2, \cdots, \mu_T, \sigma_1, \sigma_2, \cdots, \sigma_T, \rho_1 \rho_2, \cdots, \rho_{t-1}, , \beta, \theta, \psi)$, we have

$$\sqrt{n}(\hat{\eta} - \eta_0) \sim \mathbf{MVN}(\bigcirc, \mathcal{I}^{-1})$$

The Fisher information matrix $\mathcal{I}$ is estimated using the observed information matrix $\hat{\mathcal{I}}$. The Hessian matrix can be calculated during the maximization step, and the inverted Hessian matrix provides the observed Fisher information matrix.

3. Example: analysis of patient–clinician information engagement data

More and more survey studies focus on questionnaires from patients with different health issues, stages of disease, or types of cancer; both medical and more general non-medical information is needed for health providers and decision makers to better understand the behavior changes of subjects. Intuitively, patient behaviors involving attitude change, information seeking, and willingness to respond to questionnaires are related to health status. It is reasonable to expect that patients may be less likely to respond because of worsened health status; this may be a function of different kinds of health information, including disease type, patient self-efficacy, and their surrounding environment, and may contribute to informative missingness.

Table I lists the missingness patterns in PCIE data; all eight possible patterns are represented in the data, including non-monotone patterns. In practice, pattern 1, which has missing data at all three waves, carries no information and will be excluded from the study. We use the extent of exercise ('During an average week, how many days do you exercise?') as the primary outcome. The outcome ranges from 0 to 7; we treat these as continuous responses in this small interval. There were 85.66% of patients who responded to the baseline survey, 61.75% who returned the survey in wave 2, and 56.03% who answered the questions in wave 3; response rates for specific variables are given in Table II. We calculated the Pearson correlation coefficients for the exercise outcomes, which shows that the correlation is 0.644 between waves 1 and 2 and 0.618 between waves 2 and 3; the correlation between waves 1 and 3 is somewhat smaller at 0.585. We use an unstructured correlation in the data analysis and simulation sections, and we extend the correlation to AR(1), Exchangeable, and Toep(2) later in the simulation section for further model assessment.

Table III lists all patient characteristics of interest for both the marginal model and the missingness model. There are a total of 2010 cancer patients who responded to at least one of the three surveys, including 1520 patients who responded at wave 1, 1243 patients who responded at wave 2, and 1079 patients who responded at wave 3; these three surveys occurred at yearly intervals. The cohort includes both male and female patients with cancer stage from mild (stage 0) to severe (stage 4). Age at cancer diagnosis ranged from 23 to 103 years. The PCIE score was measured using eight items as described earlier, indicating whether they had

(1) Sought information about treatment from their treating physician;
(2) Sought treatment information from other physicians or health professionals;

Table I. Missingness patterns in PCIE study.

Pattern			Number of cases	Pattern			Number of cases
0	0	0	166	1	0	0	457
0	0	1	26	1	0	1	118
0	1	0	77	1	1	0	231
0	1	1	221	1	1	1	714

Table II. Response rates for possible outcomes.

Response rate	Exercise (%)	PCIE (%)	Seeking (%)
Wave 1	75.62	99.00	98.76
Wave 2	61.84	63.28	63.63
Wave 3	53.68	55.67	55.87

Table III. Patient characteristics by response time.

Response	Wave 1 ($n = 1520$) %	Wave 1 N	Wave 2 ($n = 1243$) %	Wave 2 N	Wave 3 ($n = 1079$) %	Wave 3 N
Type of cancer						
Colorectal	35.26	536	31.86	396	30.21	326
Breast	33.09	503	34.92	434	35.68	385
Prostate	31.63	481	33.23	413	34.11	368
Gender						
Male	49.28	749	48.83	607	48.29	521
Female	50.72	771	51.17	636	51.71	558
Stage						
.	6.05	92	6.11	76	6.21	67
0	9.14	139	8.21	102	10.01	108
1	17.11	260	18.99	236	19.18	207
2	38.55	586	42.24	525	44.49	480
3	12.11	184	12.15	151	11.49	124
4	17.04	259	12.31	153	8.62	93
Age	Mean	Median (range)	Mean	Median (range)	Mean	Median (range)
	64.26	65 (23–98)	63.90	64 (24–103)	63.49	64 (27–103)
PCIE score	Mean	Median (range)	Mean	Median (range)	Mean	Median (range)
	−0.006	0.0006 (−1.274 to 1.141)	−0.004	−0.212 (−0.702 to 1.886)	0.007	−0.226 (−0.604 to 2.040)

(3) Actively looked for information about their cancer from their treating physician;
(4) Looked for information about their cancer from other physicians or health professionals;
(5) Discussed information from other sources with their treating physician;
(6) Received suggestions from their treating physician to get information from other sources;
(7) Actively looked for information about quality-of-life issues from their treating physician; and
(8) Looked for quality-of-life information from other physicians or health professionals.

Each of the eight 'Yes/No' questions was answered and was transformed to a Z-score, and the average of the eight Z-scores formed the PCIE scale. The summary in Table III provides the variation in the PCIE score at each assessment time. Clearly, patients with colorectal cancer were more likely to respond at wave 1 and less likely at wave 3. Breast cancer and prostate cancer patients showed the opposite pattern. Patients with severe cancer stage were less likely to respond to the survey at wave 3 than at wave 1.

The parameters are estimated using the proposed method and compared with a generalized estimating equations (GEE) model that assumes MCAR missingness and a weighted GEE (WGEE) which assumes MAR. Both GEE and WGEE can be treated as 'ignorable' mechanisms. The original PCIE study analysis considered the missingness mechanism as either MAR or MCAR, which may have resulted in biased estimates. Our WGEE approach is a modification of the standard GEE to address missingness in the data. The response probabilities are first calculated using a logistic model for the missingness indicators; the inverse of these estimated probabilities form the corresponding weights. The missingness mechanism model used 'cancer type', 'gender', 'age at diagnosis', 'cancer severity', 'PCIE score', and the previous missingness indicator to predict current missingness indicators. For 'ignorable' data, WGEE will produce unbiased estimators if the underlying data are MAR or MCAR. GEE may have biased estimates if the underlying data are MAR.

Because missing covariate data were not of primary interest, a multiple imputation method was used to complete the missing covariates. Rubin [29] proposed a multiple imputation method using a Monte Carlo approach in which the missing values are replaced by $m > 1$ simulated versions. We generated $m = 20$ replicates in our study. Each of the imputed datasets is analyzed using both the proposed method, the GEE model, and the WGEE model. The combined parameter estimates and confidence intervals from the $m = 20$ data sets follow Rubin's [29] multiple imputation rule.

In Table IV, we list the parameter estimates after combined 20-fold imputation. The coefficient for Y_{it} indicates whether the missingness mechanism depends on the potentially unobserved outcome; a test of this parameter represents a test for non-ignorability. The coefficient for $R_{i,t-1}$ indicates whether the previous response has an effect on the likelihood of response at the current assessment; $R_{i,t-1} = 1$ indicates that the previous response was observed. Clearly, there is a statistically significant effect in the missingness model for the coefficients of both Y_{it} [−0.136 (−0.216, −0.055)] and $R_{i,t-1}$ [−0.794 (−0.955, −0.633)], which indicates that the MCAR assumption is invalid. The coefficients of Y_{it} and $R_{i,t-1}$ are both negative, indicating a negative relationship with the missingness indicator. That is, patients who exercise more days per week are less likely to have missing survey responses. They also tend to return the questionnaire if they have responded to the previous one. The coefficients for the other covariates indicated that patients who have prostate cancer [−0.302 (−0.595, −0.008)] are less likely to have missing survey responses compared with patients with colon cancer; this may reflect differences in the underlying severity of the illness. Unsurprisingly, patients with advanced cancer (stage 4) [0.713 (0.196, 1.230)] are more likely to have a missing response. 'Wave' has coefficient [0.287 (0.138, 0.424)], which suggests that patients tend to be less responsive to the survey as time passes; this is typical in repeated measures studies. Note that we assessed the missingness model for interactions between wave and the other covariates but found no such relationships.

The marginal estimates from our proposed outcome model for exercise are somewhat larger than the ones from either the GEE or WGEE approach. However, the significance levels are consistent across the models. Only 'age at diagnosis' and 'cancer stage' are statistically significant. 'Age at diagnosis' has coefficient 0.010 (0.003, 0.018), indicating that older patients engage in more exercise than younger patients do. The coefficient of 'cancer stage' [−0.567 (−1.075, −0.058)] indicates a negative correlation with outcome. Patients tend to reduce the amount of exercise when their cancer becomes more severe. PCIE did not show a statistically significant effect in either model, suggesting that in this sample, patient health behaviors are not significantly affected by patient engagement with the healthcare system as measured by the PCIE score. We assessed interaction terms in this model as well to check whether the relationship between PCIE score and exercise might be changing over time but found no evidence for this. Finally, we show the estimates in Table IV for $\rho_{1,2} = 0.643$ (0.607, 0.678) and $\rho_{2,3} = 0.624$ (0.586, 0.662); these are statistically indistinguishable, providing further justification for the choice of the first-order correlation structure.

Although the MCAR and MAR assumption is apparently invalid, both GEE and WGEE models show similar trends to the proposed model; whereas most of the parameters estimates are attenuated, the inferential conclusions are unchanged in this example. The WGEE approach provides similar results to the GEE in this example, which may be due in part to the large sample size.

4. Simulation study

4.1. Simulation results

In this section, we use a simulation study to assess model performance in small samples, addressing the basic issues of bias in the parameter estimates and computing coverage probabilities. We simulated N subjects with three potential measurement times for $N = 300$, $N = 400$, and $N = 500$. For each setting of N, we increase the missing rate from low to high. In low missingness situation, there is about 10% missing at time 1, 20–25% at time 2 , and around 40–45% missing at time 3. In the higher missingness setting, there is 30–35% missing at time 2 and 55–60% missing at time 3. The true parameters were selected to fit the proportion of each missingness pattern. The proposed method, the GEE model, and the WGEE model are applied in these six data settings. Five hundred simulations have been run to assess the model's performance; results are displayed in Table V. The data were generated as trivariate normal, with pairwise correlation parameters $\rho_{1,2} = 0.4$ and $\rho_{2,3} = 0.2$. The variance for time 1 is $\sigma_1 = 1.2$, for time 2 is $\sigma_2 = 2.6$, and for time 3 is $\sigma_3 = 3.0$. The estimators are good for both the marginal parameter

Table IV. Longitudinal analysis of PCIE data.

	Non-ignorable missing model					WGEE					GEE				
Variable	Estimate	SE	95% lower	95% upper		Estimate	SE	95% lower	95% upper		Estimate	SE	95% lower	95% upper	
Intercept	3.705	0.402	2.916	4.494	p > |z| : 0.0001	3.540	0.424	2.708	4.372	p > |z| : 0.0001	3.580	0.410	2.775	4.385	p > |z| : 0.0001
Type of cancer, breast	0.012	0.141	−0.265	0.289		0.044	0.143	−0.237	0.325		0.006	0.139	−0.266	0.278	
Type of cancer, prostate	0.189	0.155	−0.114	0.493		0.190	0.165	−0.135	0.514		0.171	0.160	−0.142	0.484	
Gender	−0.283	0.154	−0.585	0.019		−0.284	0.164	−0.606	0.038		−0.269	0.159	−0.581	0.042	
Diagnosis age	0.010	0.004	0.003	0.018	p > |z| : 0.008	0.011	0.004	0.003	0.019	p > |z| : 0.006	0.010	0.004	0.003	0.018	p > |z| : 0.043
Stage 1	−0.108	0.211	−0.522	0.306		−0.117	0.218	−0.544	0.310		−0.103	0.213	−0.521	0.315	
Stage 2	0.031	0.218	−0.396	0.458		−0.006	0.230	−0.457	0.446		0.029	0.223	−0.408	0.465	
Stage 3	−0.360	0.246	−0.842	0.122		−0.357	0.248	−0.844	0.130		−0.346	0.245	−0.826	0.134	
Stage 4	−0.567	0.259	−1.075	−0.058	p > |z| : 0.029	−0.485	0.275	−1.024	0.053	p > |z| : 0.077	−0.525	0.259	−1.033	−0.018	p > |z| : 0.008
Wave	−0.061	0.046	−0.151	0.029		0.041	0.030	−0.018	0.100		0.040	0.030	−0.018	0.098	
PCIE	−0.004	0.059	−0.120	0.112		−0.014	0.056	−0.124	0.096		−0.007	0.054	−0.112	0.098	
σ_1	2.115	0.038	2.041	2.190											
σ_2	2.116	0.043	2.032	2.199											
σ_3	2.069	0.047	1.977	2.160											
$\rho_{1,2}$	0.643	0.282	0.090	1.195											
$\rho_{2,3}$	0.624	0.273	0.090	1.158											
Missing-data model															
$y_i\,(\vartheta_c)$	−0.136	0.041	−0.216	−0.056	p > |z| : 0.0001										
$r_{i-1}\,(\psi_c)$	−0.794	0.082	−0.955	−0.633	p > |z| : 0.0001										
π_1	0.824	0.287	0.262	1.387											
Intercept	−0.491	0.419	−1.311	0.329											
Type of cancer, breast	−0.170	0.121	−0.407	0.067											
Type of cancer, prostate	−0.302	0.150	−0.595	−0.008	p > |z| : 0.044										
Gender	−0.175	0.126	−0.422	0.073											
Diagnosis age	0.011	0.004	0.004	0.019	p > |z| : 0.0018										
Stage 1	−0.208	0.249	−0.696	0.281											
Stage 2	−0.163	0.220	−0.593	0.268											
Stage 3	−0.046	0.250	−0.536	0.444											
Stage 4	0.713	0.264	0.196	1.230	p > |z| : 0.0068										
Wave	0.281	0.073	0.138	0.424	p > |z| : 0.0001										
PCIE	−0.027	0.260	−0.536	0.482											

σ_j, standard deviation of Exercise at wave j; $\rho_{j,k}$, intra-correlation of exercise at wave j and wave k.

Table V. Simulation study of sensitivity to sample size, 500 replicates.

									$N = 300$							
Response rate	Time 1 0.909	Time 2 0.840	Time 3 0.605						Time 1 0.911	Time 2 0.681	Time 3 0.434					
			Full likelihood		GEE		WGEE				Full likelihood		GEE		WGEE	
	True	E.est	Coverage probabilities	Bias	E.est	Bias	E.est	Bias	True	E.est	Coverage probabilities	Bias	E.est	Bias	E.est	Bias
Intercept	6.800	6.793	0.950	0.007	6.308	0.492	6.393	0.407	5.600	5.604	0.948	0.004	4.996	0.604	5.134	0.466
Time	1.050	1.062	0.956	0.012	1.518	0.468	1.445	0.395	0.300	0.301	0.942	0.001	0.914	0.614	0.782	0.482
Missingness model																
Intercept	−0.800	−0.990	0.958	0.190					−0.800	−0.894	0.952	0.094				
Time	1.250	1.103	0.952	0.147					1.250	1.213	0.956	0.037				
$y_i\,(\vartheta_c)$	−0.400	−0.404	0.936	0.004					−0.400	−0.412	0.950	0.012				
$r_{i-1}\,(\psi_c)$	2.500	2.676	0.976	0.176					2.500	2.628	0.962	0.128				
π	2.314	2.326	0.934	0.012					2.314	2.363	0.946	0.049				
Correlation structure																
σ_1	0.182	0.181	0.950	0.002					0.182	0.179	0.952	0.004				
σ_2	0.956	0.953	0.944	0.003					0.956	0.953	0.948	0.002				
σ_3	1.099	1.095	0.952	0.004					1.099	1.099	0.946	0.001				
$\rho_{1,2}$	0.847	0.849	0.922	0.002					0.847	0.853	0.952	0.006				
$\rho_{2,3}$	0.405	0.414	0.940	0.009					0.405	0.408	0.938	0.002				

Statistics in Medicine

Table V. *Continued*

									$N = 400$							
Response rate	Time 1 0.911	Time 2 0.826	Time 3 0.593						Time 1 0.910	Time 2 0.668	Time 3 0.425					
			Full likelihood								Full likelihood					
			Coverage		GEE		WGEE				Coverage		GEE		WGEE	
	True	E.est	probabilities	Bias	E.est	Bias	E.est	Bias	True	E.est	probabilities	Bias	E.est	Bias	E.est	Bias
Intercept	6.400	6.402	0.942	0.002	5.913	0.487	6.001	0.399	5.400	5.402	0.930	0.002	4.788	0.612	4.919	0.481
Time	1.100	1.100	0.932	0.000	1.571	0.471	1.494	0.394	0.300	0.300	0.936	0.000	0.917	0.617	0.789	0.489
Missingness model																
Intercept	−0.800	−0.867	0.954	0.067					−0.800	−0.905	0.956	0.105				
Time	1.250	1.224	0.960	0.026					1.250	1.205	0.942	0.045				
$y_i\,(\vartheta_c)$	−0.400	−0.410	0.920	0.010					−0.400	−0.407	0.940	0.007				
$r_{i-1}\,(\psi_c)$	2.500	2.600	0.950	0.100					2.500	2.612	0.948	0.112				
π	2.314	2.340	0.960	0.027					2.314	2.339	0.972	0.025				
Correlation structure																
σ_1	0.182	0.179	0.950	0.003					0.182	0.180	0.950	0.002				
σ_2	0.956	0.956	0.938	0.000					0.956	0.949	0.950	0.006				
σ_3	1.099	1.098	0.942	0.000					1.099	1.094	0.946	0.004				
$\rho_{1,2}$	0.847	0.854	0.942	0.007					0.847	0.846	0.954	0.001				
$\rho_{2,3}$	0.405	0.409	0.932	0.003					0.405	0.409	0.958	0.003				

Table V. *Continued*

									$N=500$							
Response rate	Time 1 0.911	Time 2 0.805	Time 3 0.587						Time 1 0.910	Time 2 0.666	Time 3 0.427					
			Full likelihood		GEE		WGEE				Full likelihood		GEE		WGEE	
	True	E.est	Coverage probabilities	Bias	E.est	Bias	E.est	Bias	True	E.est	Coverage probabilities	Bias	E.est	Bias	E.est	Bias
Intercept	5.600	5.591	0.940	0.009	5.111	0.489	5.202	0.398	5.400	5.404	0.952	0.004	4.787	0.613	4.918	0.482
Time	1.300	1.309	0.934	0.009	1.778	0.478	1.698	0.398	0.300	0.295	0.954	0.005	0.917	0.617	0.788	0.488
Missingness model																
Intercept	−0.800	−0.907	0.932	0.107					−0.800	−0.872	0.920	0.072				
Time	1.250	1.164	0.938	0.086					1.250	1.226	0.940	0.024				
$y_i(\vartheta_c)$	−0.400	−0.403	0.946	0.003					−0.400	−0.410	0.962	0.010				
$r_{i-1}(\psi_c)$	2.500	2.606	0.952	0.106					2.500	2.600	0.964	0.100				
π	2.314	2.334	0.938	0.020					2.314	2.347	0.936	0.033				
Correlation structure																
σ_1	0.182	0.178	0.960	0.004					0.182	0.177	0.944	0.006				
σ_2	0.956	0.956	0.944	0.000					0.956	0.957	0.946	0.001				
σ_3	1.099	1.096	0.938	0.003					1.099	1.094	0.944	0.004				
$\rho_{1,2}$	0.847	0.850	0.958	0.002					0.847	0.845	0.948	0.003				
$\rho_{2,3}$	0.405	0.405	0.950	0.001					0.405	0.417	0.948	0.011				

Simulation sample size $N = 300, 400, 500$.

σ_j, standard deviation of outcome at wave j; $\rho_{j,k}$, intra-correlation of outcome at waves j and k.

Statistics in Medicine

Table VI. Simulation study of model comparison, 1000 replicates.

	$N=300$									
Response rate time 1		0.909					0.910			
Response rate time 2		0.840					0.881			
Response rate time 3		0.604					0.637			
		Transition model correctly specified		Troxel 1998 model			Transition model correctly specified		Troxel 1998 model	
	True	E.est	CP	E.est	CP	True	E.est	CP	E.est	CP
Intercept	6.8	6.798	0.944	6.711	0.926	1.8	1.788	0.931	1.780	0.932
time	1.05	1.048	0.935	1.133	0.911	1.05	1.062	0.931	1.070	0.933
Missingness model										
Intercept	−0.8	−0.948	0.954	0.721	0.575	−0.8	−1.115	0.941	−0.919	0.958
Time	1.25	1.145	0.943	2.397	0.852	1.25	1.022	0.944	1.190	0.949
$y_i(\vartheta_c)$	−0.4	−0.410	0.932	−0.289	0.785	−0.4	−0.407	0.950	−0.396	0.942
$r_{i-1}(\psi_c)$	2.5	2.674	0.963	–	–	0	0.229	0.959	–	–
π	0.91	0.910	1.000	0.910	1.000	0.91	0.914	1.000	0.916	1.000
Correlation										
σ_1	1.2	1.195	0.967	1.196	0.968	1.2	1.195	0.940	1.194	0.940
σ_2	2.6	2.593	0.948	2.549	0.919	2.6	2.594	0.936	2.586	0.936
σ_3	3	3.001	0.940	2.949	0.932	3	2.997	0.928	2.978	0.929
$\rho_{1,2}$	0.4	0.398	0.953	0.400	0.950	0.4	0.400	0.948	0.402	0.948
$\rho_{2,3}$	0.2	0.202	0.948	0.207	0.942	0.2	0.196	0.940	0.196	0.940

Simulation sample size $N=300$.

CP, coverage probability; σ_j, standard deviation of outcome at wave j; $\rho_{j,k}$, intra-correlation of outcome at wave j and wave k.

and the missingness mechanism model in Table V. The bias is very small. Both GEE and WGEE methods consistently underestimate the parameters when the sample size is small and the bias is substantial. This becomes much more severe when the missingness rate increases from low to high. For WGEE, the estimated response probabilities are obtained through a logistic regression first, and the inverse of these observed probabilities form the weights. 'Wave' and previous response indicator $R_{i,t-1}$ are used to predict the observed probabilities. As expected, WGEE provides better estimators than GEE across all data settings, especially when the proportion of missing data is high, although the bias is still substantial compared with the proposed approach.

4.2. *Model comparison*

The proposed model is compared with the original model in Troxel *et al.* [9], which used the same setting for the complete data and a different logistic model for the missingness indicators denoted as $\text{logit}(\pi_{r_{it}=1}) = \beta_{0t} + \beta_1 Y_{it}$. In the Troxel model [9], this missingness model did not include the previous missing indicator as a covariate. We generated two data settings, one with our proposed model and one with the correctly specified Troxel model [9]. The correctly specified original model from Troxel *et al.* [9] will become a mis-specified model if the coefficient of the previous missing indicator is not zero. Our proposed model will be over-specified if the parameter of the previous missing indicator is zero. Table VI shows these comparison results. When the parameter (ψ_c) of the previous missing indicator is not zero, the estimates from our transition model are unbiased and have high coverage probabilities. The Troxel 1998 model has good estimation in the marginal model and variance–covariance structure but poor estimation in the missingness model. This is not surprising, as the missingness model is mis-specified. When the parameter (ψ_c) of the previous missing indicator is zero, both models have very good estimation. The proposed model uses a small value to estimate the ψ_c with 95% coverage rate including zero.

Table VII. Simulation study of sensitivity to different covariance structures, 500 replicates.

		AR(1) $\sigma^2\rho^{i-j}$		Exchangeable $\sigma^2[\rho 1(i \neq j) + 1(i = j)]$		Toep(2) $\sigma^2_{\vert i-j\vert+1} 1(\vert i-j\vert < 2)$	
Response rate time 1		0.910		0.910		0.910	
Response rate time 2		0.844		0.844		0.844	
Response rate time 3		0.614		0.616		0.610	
	True	E.est	CP	E.est	CP	E.est	CP
Intercept	6.8	6.809	0.954	6.799	0.954	6.809	0.932
Time	1.05	1.047	0.952	1.055	0.970	1.037	0.920
Missing-data model							
Intercept	−0.8	−0.841	0.962	−0.934	0.960	−0.749	0.944
Time	1.25	1.238	0.970	1.123	0.954	1.331	0.966
$y_i(\vartheta_c)$	−0.4	−0.408	0.962	−0.397	0.952	−0.423	0.948
$r_{i-1}(\psi_c)$	2.5	2.581	0.966	2.593	0.968	2.622	0.948
π	0.904	0.911	1.000	0.911	1.000	0.910	1.000
correlation							
σ_1	2.4	2.404	0.946	2.401	0.946	2.403	0.960
σ_2	2.4	2.407	0.956	2.421	0.944	2.385	0.956
σ_3	2.4	2.405	0.930	2.405	0.956	2.403	0.956
$\rho_{1,2}$	0.6	0.600	0.954	0.613	0.932	0.583	0.920
$\rho_{2,3}$	0.6	0.600	0.962	0.620	0.894	0.568	0.876

Simulation sample size $N = 500$.
CP, coverage probability; σ_j, standard deviation of outcome at wave j; $\rho_{j,k}$, correlation of outcome at wave j and wave k.

Statistics in Medicine

Next, we fit the proposed model with three different covariance structures to see how our model handles a mis-specified correlation matrix. Our transition model uses ANTE(1) (ante-dependence) structure denoted as $\sigma_i \sigma_j \prod_{k=i}^{j-1} \rho_k$ for the (i, j)th element. There are a total of $2t-1$ parameters needed. This will become computationally burdensome when t, the number of repeated times, increases. In practice, the AR(1) (autoregressive(1)) structure is widely used, denoted as $\sigma^2 \rho^{i-j}$ for the (i, j)th element. There are only two parameters needed. Another two correlation structures used for comparison are Exchangeable $(\sigma^2[\rho 1(i \neq j) + 1(i = j)])$ structure and Toep(2) (Banded Toeplitz $\sigma^2_{|i-j|+1} 1(|i-j| < 2)$) structure.

The AR(1), Exchangeable, and Toep(2) structures for $T = 3$ are written respectively as follows:

$$\Sigma = \begin{Bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{Bmatrix}_{\mathrm{AR(1)}} ; \Sigma = \begin{Bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2\rho & \sigma^2 \end{Bmatrix}_{\mathrm{Exch}} ; \Sigma = \begin{Bmatrix} \sigma^2 & \sigma_1 & 0 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ 0 & \sigma_1 & \sigma^2 \end{Bmatrix}_{\mathrm{Toep(2)}}$$

Table VIII. Simulation study of normal data versus gamma data, 500 replicates.

				$\rho = 0.707$				
		Normal				Gamma		
Response rate	0.910	0.855	0.613		0.910	0.855	0.609	
	True	Mean Estimation	Standard Deviation	CP	True	Mean Estimation	Standard Deviation	CP
Intercept	6.800	6.795	0.142	0.958	6.800	6.846	0.155	0.950
Time	1.050	1.054	0.065	0.958	1.050	1.000	0.080	0.912
Intercept	−0.800	−1.002	0.992	0.938	−0.800	−0.726	0.890	0.932
Time	1.250	1.055	1.063	0.946	1.250	1.417	0.988	0.916
$y_i(\vartheta_c)$	−0.400	−0.398	0.078	0.934	−0.400	−0.450	0.097	0.892
$r_{i-1}(\psi_c)$	2.500	2.662	0.805	0.972	2.500	2.760	0.640	0.972
π	2.314	2.337	0.205	0.930	2.314	2.336	0.205	0.956
Correlation structure								
σ_1	1.772	1.764	0.075	0.940	1.772	1.765	0.075	0.922
σ_2	1.887	1.898	0.083	0.962	1.887	1.957	0.088	0.846
σ_3	1.995	1.988	0.099	0.938	1.995	2.109	0.111	0.800
$\rho_{1,2}$	0.707	0.716	0.079	0.930	0.707	0.637	0.078	0.470
$\rho_{2,3}$	0.707	0.723	0.093	0.918	0.707	0.603	0.093	0.324
				$\rho = 0.5$				
		Normal				Gamma		
Response rate	0.909	0.857	0.609		0.910	0.856	0.607	
	True	E.est	STD	CP	True	E.est	STD	CP
Intercept	6.800	6.796	0.165	0.962	6.800	6.877	0.181	0.918
Time	1.050	1.056	0.086	0.970	1.050	0.967	0.106	0.866
Intercept	−0.800	−1.178	1.293	0.954	−0.800	−0.291	1.150	0.876
Time	1.250	0.876	1.395	0.942	1.250	1.964	1.332	0.876
$y_i(\vartheta_c)$	−0.400	−0.385	0.097	0.934	−0.400	−0.527	0.153	0.862
$r_{i-1}(\psi_c)$	2.500	2.724	1.010	0.988	2.500	2.868	0.696	0.982
π	2.314	2.323	0.204	0.944	2.314	2.338	0.205	0.932
Correlation structure								
σ_1	1.772	0.572	0.075	0.958	1.772	1.763	0.075	0.924
σ_2	1.887	0.635	0.086	0.958	1.887	1.966	0.094	0.840
σ_3	1.995	0.691	0.106	0.954	1.995	2.145	0.127	0.762
$\rho_{1,2}$	0.500	1.099	0.076	0.948	0.500	0.454	0.076	0.820
$\rho_{2,3}$	0.500	1.099	0.092	0.932	0.500	0.410	0.095	0.704

Table VIII. *Continued*

	$\rho = 0$							
	Normal				Gamma			
Response rate	0.910	0.857	0.600		0.910	0.857	0.600	
	True	E.est	STD	CP	True	E.est	STD	CP
Intercept	6.800	6.784	0.228	0.934	6.800	6.996	0.229	0.710
Time	1.050	1.064	0.161	0.892	1.050	0.858	0.159	0.498
Intercept	−0.800	−1.299	2.297	0.896	−0.800	1.267	2.702	0.538
Time	1.250	0.763	2.609	0.894	1.250	3.959	3.082	0.528
$y_i\,(\vartheta_c)$	−0.400	−0.392	0.234	0.876	−0.400	−0.820	0.294	0.490
$r_{i-1}\,(\psi_c)$	2.500	2.821	1.144	0.968	2.500	3.137	1.552	0.986
π	2.314	2.329	0.204	0.934	2.314	2.335	0.205	0.934
Correlation structure								
σ_1	1.772	1.767	0.076	0.958	1.772	1.761	0.075	0.934
σ_2	1.887	1.896	0.103	0.926	1.887	2.033	0.116	0.648
σ_3	1.995	2.000	0.144	0.940	1.995	2.294	0.182	0.562
$\rho_{1,2}$	0.000	−0.002	0.066	0.946	0.000	−0.004	0.064	0.942
$\rho_{2,3}$	0.000	0.004	0.081	0.932	0.000	−0.026	0.074	0.888

Simulation sample size $N = 300$.
CP, coverage probability; σ_j, standard deviation of outcome at wave j; $\rho_{j,k}$, correlation of outcome at wave j and wave k.

The comparison table is listed in Table VII. The proposed model can handle the AR(1) structure well as it is a special case of ANTE(1) structure. Our model still performs quite well in estimating the marginal effects and missingness coefficients for both the Exchangeable and Toep(2) structures. The variances are estimated with high coverage probabilities. Both correlation estimates are less efficient than for the AR(1) model.

4.3. Non-normal data

In this section, we compare the proposed models in cases with different underlying assumptions about the true data distribution. We simulated two data sets with same expected values but with different distributions. One scenario was simulated from a trivariate normal distribution. A second scenario was simulated using a trivariate gamma distribution. A Clayton copula, which is an asymmetric Archimedean copula, was used to generate the trivariate gamma data. The dependence structure of the trivariate gamma followed an Exchangeable correlation structure. We used Kendall's formula [30] to assure the same covariance structure between the trivariate normal and trivariate gamma data. We generated three correlation structures, with high ($\rho = 0.707$), low ($\rho = 0.5$), and zero ($\rho = 0$) pairwise correlation, with sample size $n = 300$ and $n = 500$ replications to examine the model's performance. The comparison table is listed in Table VIII. Our proposed model performed quite well even with the mis-specified distribution. The estimator becomes less efficient when the assessments are independent ($\rho = 0$); this is not surprising as in this scenario the missingness model is mis-specified and both models are over-fitted. The marginal effect and missingness models are still estimated well when the underlying data distribution is not normal. The correlation coefficients, however, are poorly estimated. The estimation improves when the correlation is strong and worsens when the correlation is weak in the dependent data cases.

5. Discussion

We have presented an extension of the full likelihood-based algorithm to handle non-monotone and non-ignorable missing data. We assume a first-order Markov structure in both the complete data and missingness mechanism which is a natural way to capture the correlation among repeated measurements in a longitudinal data framework. The estimation of marginal effects is generally robust to correct specification of the covariance matrix and missingness mechanism.

As with any model-based approach to non-ignorable missing data, the current approach is subject to unavoidable assumptions about the complete data distribution and the missing-data mechanism. It is important to consider all substantive information about the area of application, prior experience with missing data in similar situations, and expert opinion about the mechanism of missing data when building such models. In many areas, enough knowledge and experience exists to justify the necessary assumptions, and the benefit in terms of bias reduction can be significant.

Our transition model can be easily extended to model more than two states such as dropout or intermittent missingness. The numerical integration provides an accurate approximation but at the cost of increased computational complexity. We occasionally encountered a multimodal likelihood surface in our study. A method to handle such surface is to choose a vector of starting values by using GEE estimates to get the starting point as close to the true values as possible. There are too many classes of correlation structure to explore them all; however, the proposed model can handle a mis-specified correlation to some extent. In simulation studies with a variety of mis-specified correlation structures, the marginal effects and missingness effects consistently have high coverage probabilities as long correlation among pairs is nonzero. For vectors of observation times larger than three, it will be difficult to examine many correlation structures because of the increase in the number of parameters in the model.

Given the increasing interest in healthcare reform and structural changes in healthcare systems, more and more survey studies are being designed to better understand the relationships among patients, physicians, and the broader healthcare system. In many such studies, however, sample sizes are limited based on the disease under study, the geographic area, and the availability of medical information. Small sample sizes with a large proportion of missing information become a vexing problem for researchers trying to evaluate the associations of interest. The missingness probability is often related to the very outcomes under study, for example, when patients fail to respond because of worse health outcomes. In the example studied here, the level of exercise may well serve as a proxy for general health status; ignoring this information in the analysis can lead to seriously biased results.

APPENDIX A. Conditional density

For $T = 3$, assume that the first observation depends on other covariates and is always observed. For $T = 1$ then

$$
\begin{aligned}
f(y_{i1}) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right) \\
f(R_{i1}|y_{i1}) &= \pi_{i1}^{R_{i1}}(1-\pi_{i1})^{1-R_{i1}}
\end{aligned}
$$

For $T = 2$ then

$$
\begin{aligned}
f(y_{i2}|y_{i1}) &= \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \\
f(R_{i2}=1|R_{i1},y_{i2}) &= \frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}}+\psi_1\mathbf{R_{i1}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}}+\psi_s\mathbf{R_{i1}})}.
\end{aligned}
$$

For $T = 3$ then

$$
\begin{aligned}
f(y_{i3}|y_{i2}) &= \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \exp\left(\frac{1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
f(R_{i3}=1|R_{i2},y_{i3}) &= \frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1\mathbf{Y_{i3}}+\psi_1\mathbf{R_{i2}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}}+\psi_s\mathbf{R_{i2}})}.
\end{aligned}
$$

APPENDIX B. Joint likelihood function

For $T = 3$ there are $2^3 = 8$ patterns. If we do not allow all points to be missing $\begin{pmatrix} * & * & * \\ R_{i1} & R_{i2} & R_{i3} \\ 0 & 0 & 0 \end{pmatrix}$ then we will have seven patterns. The possible patterns are listed below.

Pattern 1 $P_1 := \begin{pmatrix} Y_{i1} & Y_{i2} & Y_{i3} \\ R_{i1} & R_{i2} & R_{i3} \\ 1 & 1 & 1 \end{pmatrix}$

$$
\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p1} =& \mathcal{L}_i \\
=& f(y_{i1}) f(y_{i2}|y_{i1}) f(y_{i3}|y_{i2}) f(R_{i1}|y_{i1}) f(R_{i2}|R_{i1}, y_{i2}) f(R_{i3}|R_{i2}, y_{i3}) \\
=& \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1} \\
&\times \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \\
&\times \frac{\exp(\zeta_1'\mathbf{X_{i2}} + \vartheta_1\mathbf{Y_{i2}} + \psi_1)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i2}} + \vartheta_s\mathbf{Y_{i2}} + \psi_s)} \\
&\times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
&\times \frac{\exp(\zeta_1'\mathbf{X_{i3}} + \vartheta_1 Y_{i3} + \psi_1)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i3}} + \vartheta_s Y_{i3} + \psi_s)}
\end{aligned}
$$

Pattern 2 $P_2 := \begin{pmatrix} Y_{i1} & Y_{i2} & * \\ R_{i1} & R_{i2} & R_{i3} \\ 1 & 1 & 0 \end{pmatrix}$

$$
\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p2} =& \int \mathcal{L}_i \, \mathrm{d}y_{i3} \\
=& \int f(y_{i1}) f(y_{i2}|y_{i1}) f(y_{i3}|y_{i2}) f(R_{i1}|y_{i1}) f(R_{i2}|R_{i1}, y_{i2}) f(R_{i3}|R_{i2}, y_{i3}) \mathrm{d}y_{i3} \\
=& \int f(R_{i3}|R_{i2}, y_{i3}) f(y_{i3}|y_{i2}) \mathrm{d}y_{i3} \\
&\times f(y_{i1}) f(y_{i2}|y_{i1}) f(R_{i1}|y_{i1}) f(R_{i2}|R_{i1}, y_{i2}) \\
=& \mathbb{E}_{f_{3|2}}(f(R_{i3}|R_{i2}, y_{i3}^*)) \times f(y_{i1}) f(y_{i2}|y_{i1}) f(R_{i1}|y_{i1}) f(R_{i2}|R_{i1}, y_{i2}) \\
=& \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1} \\
&\times \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \\
&\times \frac{\exp(\zeta_1'\mathbf{X_{i2}} + \vartheta_1\mathbf{Y_{i2}} + \psi_1)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i2}} + \vartheta_s\mathbf{Y_{i2}} + \psi_s)} \\
&\times \int \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
&\times \frac{\exp(\zeta_2'\mathbf{X_{i3}} + \vartheta_2\mathbf{Y_{i3}} + \psi_2)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i3}} + \vartheta_s Y_{i3} + \psi_s)} \mathrm{d}y_{i3} \\
=& \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_{i1}^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1}
\end{aligned}
$$

$$\times \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right)$$
$$\times \frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}}+\psi_1)}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}}+\psi_s)}$$
$$\times \sum_{k=1}^m \frac{w_k}{\sqrt{\pi}} \frac{\exp(\zeta_2'\mathbf{X_{i3}}+\vartheta_2(\mu_{i3}+\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})+\sqrt{2\sigma_3^2(1-\rho_2^2)}\tau_k)+\psi_2)}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s(\mu_{i3}+\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})+\sqrt{2\sigma_3^2(1-\rho_2^2)}\tau_k)+\psi_s)}$$

Pattern 3 $P_3 := \begin{pmatrix} Y_{i1} & * & * \\ R_{i1} & R_{i2} & R_{i3} \\ 1 & 0 & 0 \end{pmatrix}$

$$\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p3} &= \int\int \mathcal{L}_i \mathrm{d}y_{i2}\mathrm{d}y_{i3} \\
&= \int\int f(R_{i3}|R_{i2},y_{i3})f(R_{i2}|R_{i1},y_{i2})f(y_{i3}|y_{i2})f(y_{i2}|y_{i1})\mathrm{d}y_{i2}\mathrm{d}y_{i3}f(R_{i1}|y_{i1})f(y_{i1}) \\
&= \int\int f(R_{i3}|R_{i2},y_{i3})f(R_{i2}|R_{i1},y_{i2})f(y_{i3},y_{i2}|y_{i1})\mathrm{d}y_{i2}\mathrm{d}y_{i3}f(R_{i1}|y_{i1})f(y_{i1}) \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(\frac{-1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1} \\
&\quad \times \int\int \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}}\exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \\
&\quad \times \frac{\exp(\zeta_2'\mathbf{X_{i2}}+\vartheta_2Y_{i2}+\psi_2)}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_sY_{i2}+\psi_s)} \\
&\quad \times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}\exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
&\quad \times \frac{\exp(\zeta_2'\mathbf{X_{i3}}+\vartheta_2\mathbf{Y_{i3}})}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}})}\mathrm{d}y_{i2}\mathrm{d}y_{i3} \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(\frac{-1}{2\sigma_{i1}^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1} \\
&\quad \times \int\sum_{k=1}^m\frac{w_k}{\sqrt{\pi}}\frac{\exp(\zeta_2'\mathbf{X_{i2}}+\vartheta_2(\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k)+\psi_2)}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s(\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k)+\psi_s)} \\
&\quad \times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \\
&\quad \times \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}\left(\left(\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k\right)-\mu_{i2}\right)\right)^2\right) \\
&\quad \times \frac{\exp(\zeta_2'\mathbf{X_{i3}}+\vartheta_2\mathbf{Y_{i3}})}{\sum_{s=1}^2\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}})}\mathrm{d}y_{i3} \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(\frac{-1}{2\sigma_{i1}^2}(y_{i1}-\mu_{i1})^2\right)\pi_{i1}
\end{aligned}$$

$$\times \sum_{k=1}^{m} \frac{w_k}{\sqrt{\pi}} \frac{\exp(\zeta_2' \mathbf{X_{i2}} + \vartheta_2(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \psi_2)}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i2}} + \vartheta_s(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \psi_s)}$$

$$\times \int \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}$$

$$\times \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3} - \mu_{i3} - \rho_2 \frac{\sigma_3}{\sigma_2}\left(\rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k\right)\right)^2\right)$$

$$\times \frac{\exp\left(\zeta_2' \mathbf{X_{i3}} + \vartheta_2 \mathbf{Y_{i3}}\right)}{\sum_{s=1}^{2} \exp\left(\zeta_s' \mathbf{X_{i3}} + \vartheta_s \mathbf{Y_{i3}}\right)} \mathrm{d}y_{i3}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1} - \mu_{i1})^2\right) \pi_{i1}$$

$$\times \sum_{k=1}^{m} \frac{w_k}{\sqrt{\pi}} \frac{\exp\left(\zeta_2' \mathbf{X_{i2}} + \vartheta_2(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \psi_2\right)}{\sum_{s=1}^{2} \exp\left(\zeta_s' \mathbf{X_{i2}} + \vartheta_s\left(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k\right) + \psi_s\right)}$$

$$\times \sum_{l=1}^{m} \frac{w_l}{\sqrt{\pi}} \frac{\exp(\zeta_2' \mathbf{X_{i3}} + \vartheta_2(\mu_{i3} + \rho_2 \frac{\sigma_3}{\sigma_2}(\rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \sqrt{2\pi\sigma_3^2(1-\rho_2^2)}\tau_l))}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i3}} + \vartheta_s(\mu_{i3} + \rho_2 \frac{\sigma_3}{\sigma_2}(\rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \sqrt{2\pi\sigma_3^2(1-\rho_2^2)}\tau_l))}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1} - \mu_{i1})^2\right) \pi_{i1}$$

$$\times \sum_{k=1}^{m} \sum_{l=1}^{m} \frac{w_k w_l}{\pi} \frac{\exp(\zeta_2' \mathbf{X_{i2}} + \vartheta_2(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \psi_2)}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i2}} + \vartheta_s(\mu_{i2} + \rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \psi_s)}$$

$$\times \frac{\exp(\zeta_2' \mathbf{X_{i3}} + \vartheta_2(\mu_{i3} + \rho_2 \frac{\sigma_3}{\sigma_2}(\rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \sqrt{2\pi\sigma_3^2(1-\rho_2^2)}\tau_l))}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i3}} + \vartheta_s(\mu_{i3} + \rho_2 \frac{\sigma_3}{\sigma_2}(\rho_1 \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}) + \sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k) + \sqrt{2\pi\sigma_3^2(1-\rho_2^2)}\tau_l))}$$

$$\text{Pattern 4 } P_4 := \begin{pmatrix} Y_{i1} & * & Y_{i3} \\ R_{i1} & R_{i2} & R_{i3} \\ 1 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p4} &= \int \mathcal{L}_i \mathrm{d}y_{i2} \\
&= \int f(y_{i1}) f(y_{i2}|y_{i1}) f(y_{i3}|y_{i2}) f(R_{i1}|y_{i1}) f(R_{i2}|R_{i1}, y_{i2}) f(R_{i3}|R_{i2}, y_{i3}) \mathrm{d}y_{i2} \\
&= \int f(R_{i2}|R_{i1}, y_{i2}) f(y_{i3}|y_{i2}) f(y_{i2}|y_{i1}) \mathrm{d}y_{i2} \\
&\quad \times f(y_{i1}) f(R_{i1}|y_{i1}) f(R_{i3}|R_{i2}, y_{i3}) \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1} - \mu_{i1})^2\right) \pi_{i1} \frac{\exp(\zeta_1' \mathbf{X_{i3}} + \vartheta_1 \mathbf{Y_{i3}})}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i3}} + \vartheta_s \mathbf{Y_{i3}})} \\
&\quad \times \int \frac{\exp(\zeta_2' \mathbf{X_{i2}} + \vartheta_2 \mathbf{Y_{i2}} + \psi_2)}{\sum_{s=1}^{2} \exp(\zeta_s' \mathbf{X_{i2}} + \vartheta_s \mathbf{Y_{i2}} + \psi_s)} \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \\
&\quad \times \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3} - \mu_{i3} - \rho_2 \frac{\sigma_3}{\sigma_2}(y_{i2} - \mu_{i2})\right)^2\right)
\end{aligned}$$

$$\times \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \mathrm{d}y_{i2}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right) \pi_{i1} \frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1\mathbf{Y_{i3}})}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}})}$$

$$\times \sum_{k=1}^{m} \frac{w_k}{\sqrt{\pi}} \frac{\exp(\zeta_2'\mathbf{X_{i2}}+\vartheta_2(\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k)+\psi_2)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s(\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k)+\psi_s)}$$

$$\times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}$$

$$\times \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}((\mu_{i2}+\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_k)-\mu_{i2})\right)^2\right)$$

Pattern 5 $P_5 := \begin{pmatrix} * & Y_{i2} & Y_{i3} \\ R_{i1} & R_{i2} & R_{i3} \\ 0 & 1 & 1 \end{pmatrix}$

$$\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p5} &= \int \mathcal{L}_i \mathrm{d}y_{i1} \\
&= \int f(R_{i1}|y_{i1}) f(y_{i2}|y_{i1}) f(y_{i1}) \mathrm{d}y_{i1} f(R_{i3}|R_{i2},y_{i3}) f(R_{i2}|R_{i1},y_{i2}) f(y_{i3}|y_{i2}) \\
&= \frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1 Y_{i3}+\psi_1)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s Y_{i3}+\psi_s)} \frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}})}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}})} \\
&\quad \times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
&\quad \times (1-\pi_{i1}) \int \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right) \frac{1}{\sqrt{2\pi\sigma_1^2}} \\
&\quad \times \exp\left(\frac{1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right) \mathrm{d}y_{i1} \\
&= (1-\pi_{i1}) \frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1\mathbf{Y_{i3}}+\psi_1)}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}}+\psi_s)} \frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}})}{\sum_{s=1}^2 \exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}})} \\
&\quad \times \frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}} \exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right) \\
&\quad \times \sum_{k=1}^{m} \frac{w_k}{\sqrt{\pi}} \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}} \exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}((\mu_{i1}+\sqrt{2\sigma_1^2}\tau_k)-\mu_{i1})\right)^2\right)
\end{aligned}$$

Pattern 6 $P_6 := \begin{pmatrix} * & Y_{i2} & * \\ R_{i1} & R_{i2} & R_{i3} \\ 0 & 1 & 0 \end{pmatrix}$

$$\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p6} &= \int\int \mathcal{L}_i \mathrm{d}y_{i1}\mathrm{d}y_{i3} \\
&= \int\int f(y_{i1}) f(y_{i2}|y_{i1}) f(y_{i3}|y_{i2}) f(R_{i1}|y_{i1}) f(R_{i3}|R_{i2},y_{i3}) \mathrm{d}y_{i1}\mathrm{d}y_{i3} f(R_{i2}|R_{i1},y_{i2}) \\
&= \int f(R_{i3}|R_{i2},y_{i3}) f(y_{i3}|y_{i2}) \mathrm{d}y_{i3} \times \int f(y_{i1}) f(y_{i2}|y_{i1}) f(R_{i1}|y_{i1}) \mathrm{d}y_{i1} \times f(R_{i2}|R_{i1},y_{i2})
\end{aligned}$$

$$
\begin{aligned}
&=(1-\pi_{i1})\frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}})}\\
&\quad\times\int\frac{\exp(\zeta_2'\mathbf{X_{i3}}+\vartheta_2\mathbf{Y_{i3}}+\psi_2)}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}}+\psi_s)}\frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}\\
&\quad\times\exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right)\mathrm{d}y_{i3}\\
&\quad\times\int\frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}}\exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right)\frac{1}{\sqrt{2\pi\sigma_1^2}}\\
&\quad\times\exp\left(\frac{1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right)\mathrm{d}y_{i1}\\
&=(1-\pi_{i1})\frac{\exp(\zeta_1'\mathbf{X_{i2}}+\vartheta_1\mathbf{Y_{i2}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}})}\\
&\quad\times\sum_{k=1}^{m}\sum_{l=1}^{m}\frac{w_kw_l}{\pi}\frac{\exp(\zeta_2'\mathbf{X_{i3}}+\vartheta_2(\mu_{i3}+\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})+\sqrt{2\sigma_3^2(1-\rho_2^2)}\tau_k)+\psi_2)}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s*(\mu_{i3}+\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})+\sqrt{2\sigma_3^2(1-\rho_2^2)}\tau_k)+\psi_s)}\\
&\quad\times\frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}}\exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}(y_{i2}-\mu_{i2}-\sqrt{2}\rho_1\sigma_2\tau_l)^2\right)
\end{aligned}
$$

Pattern 7 $P_7:=\begin{pmatrix} * & * & Y_{i3}\\ R_{i1} & R_{i2} & R_{i3}\\ 0 & 0 & 1\end{pmatrix}$

$$
\begin{aligned}
\mathcal{L}_{i,\text{obs}}^{p7}&=\int\int\mathcal{L}_i\,\mathrm{d}y_{i1}\mathrm{d}y_{i2}\\
&=\int\int f(R_{i1}|y_{i1})f(R_{i2}|R_{i1},y_{i2})f(y_{i1}f(y_{i2}|y_{i1})f(y_{i3}|y_{i2})\mathrm{d}y_{i1}\mathrm{d}y_{i2}f(R_{i3}|R_{i2},y_{i3})\\
&=\frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1\mathbf{Y_{i3}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}})}(1-\pi_{i1})\int\int\frac{\exp(\zeta_2'\mathbf{X_{i2}}+\vartheta_2\mathbf{Y_{i2}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s\mathbf{Y_{i2}})}\frac{1}{\sqrt{2\pi\sigma_1^2}}\\
&\quad\times\exp\left(\frac{-1}{2\sigma_1^2}(y_{i1}-\mu_{i1})^2\right)\\
&\quad\times\frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho_1^2)}}\exp\left(\frac{-1}{2\sigma_2^2(1-\rho_1^2)}\left(y_{i2}-\mu_{i2}-\rho_1\frac{\sigma_2}{\sigma_1}(y_{i1}-\mu_{i1})\right)^2\right)\\
&\quad\times\frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}\exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(y_{i2}-\mu_{i2})\right)^2\right)\mathrm{d}y_{i1}\mathrm{d}y_{i2}\\
&=\frac{\exp(\zeta_1'\mathbf{X_{i3}}+\vartheta_1\mathbf{Y_{i3}})}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i3}}+\vartheta_s\mathbf{Y_{i3}})}(1-\pi_{i1})\\
&\quad\times\sum_{k=1}^{m}\sum_{l=1}^{m}\frac{w_kw_l}{\pi}\frac{\exp(\zeta_2'\mathbf{X_{i2}}+\vartheta_2(\mu_{i2}+\sqrt{2}\rho_1\sigma_2\tau_k+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_l))}{\sum_{s=1}^{2}\exp(\zeta_s'\mathbf{X_{i2}}+\vartheta_s(\mu_{i2}+\sqrt{2}\rho_1\sigma_2\tau_k+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_l))}\\
&\quad\times\frac{1}{\sqrt{2\pi\sigma_3^2(1-\rho_2^2)}}\\
&\quad\times\exp\left(\frac{-1}{2\sigma_3^2(1-\rho_2^2)}\left(y_{i3}-\mu_{i3}-\rho_2\frac{\sigma_3}{\sigma_2}(\mu_{i2}+\sqrt{2}\rho_1\sigma_2\tau_k+\sqrt{2\sigma_2^2(1-\rho_1^2)}\tau_l-\mu_{i2})\right)^2\right)
\end{aligned}
$$

Statistics
in Medicine

Acknowledgements

The authors are grateful to Robert C. Hornik, Katrina Armstrong, Mary E. Putt, and Benjamin French for their insightful comments on earlier drafts. Kaijun Liao gratefully acknowledges support from the Annenberg School for Communication, University of Pennsylvania and support from the National Cancer Institute (grant # P50-CA095856-05).

References

1. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*, 2nd Edition. Wiley: New York, 1987.
2. Allison PD. *Missing Data*, Thousand Oaks. Sage Publications, Inc: CA, 2001.
3. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
4. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.
5. Verbyla AP, Cullis BR. Modelling in repeated measures experiments. *Journal of the Royal Statistical Society* 1990; **39**:341–356.
6. Richard H, Lynn M. Serial correlation in unequally spaced longitudinal data. *Biometrika* 1990; **77**(4):721–732.
7. Munoz A, Carey V, Schouten J, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 1992; **48**:733–742.
8. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994; **43**:49–93.
9. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society* 1998; **47**:425–438.
10. Troxel AB, Lipsitz SR, Harrington DP. Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* 1998; **85.3**:661–672.
11. Sinha SK, Troxel AB, Lipsitz SR, Sinha D, Fitzmaurice GM, Molenberghs G, Ibrahim JG. A bivariate pseudolikelihood method for incomplete longitudinal binary data with nonignorable non-monotone missingness. *Biometrics* 2010; **25**:2784–2789.
12. Parzen M, Lipsitz SR, Fitzmaurice GM, Ibrahim JG, Troxel AB, Molenberghs G. Pseudo-likelihood methods for the analysis of longitudinal binary data subject to nonignorable non-monotone missing. *Journal of data science* 2007; **5**:1–21.
13. Shen C, Weissfeld L. A copula model for repeated measurements with non-ignorable non-monotone missing outcome. *Statistics in Medicine* 2006; **25**:2427–2440.
14. Troxel AB, Lipsitz SR, Fitzmaurice GM, Ibrahim JG, Sinhad D, Molenberghse G. A weighted combination of pseudo-likelihood estimators for longitudinal binary data subject to non-ignorable non-monotone missingness. *Statistics in Medicine* 2010; **29**:1511-1521.
15. Shardell M, Miller RR. Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine* 2008; **27**:1008–1025.
16. Birmingham J, Fitzmaurice GM. A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics* 2002; **58**:989–996.
17. Pauler DK, McCoy S, Moinpour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine* 2003; **22**:795-809. DOI: 10.1002/sim.1397.
18. Shardell M, Hicks GE, Miller RR, Langenberg P, Magaziner J. Pattern-mixture models for analyzing normal outcome data with proxy respondents. *Statistics in Medicine* 2010; **29**(14):1522-1538.
19. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics* 2009; **65**:81–87.
20. Robins JM. Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine* 1997; **16**:21–37.
21. Robins JM, Gill R. Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine* 1997; **16**:39–56.
22. Liu Q, Pierce DA. A note on Gauss Hermite quadrature. *Biometrika* 1994; **81**:624-629.
23. Nocedal J, Wright SJ. *Numerical Optimization*, 2nd Edition: Berlin, New York, 2006.
24. Putt M, Long JA, Montagnet C, Silber JH, Chang VW, Liao K, Schwartz JS, Pollack CE, Wong YN, Armstrong K. Racial differences in the impact of comorbidities on survival among elderly men with prostate cancer. *Medical Care Research Review* 2009; **66**:409-35.
25. Tan AS, Bourgoin A, Gray SW, Armstrong K, Hornik RC. How does patient–clinician information engagement influence self-reported cancer-related problems?. *Cancer* 2011; **117**(11):2569–2576. DOI: 10.1002/cncr.25804.
26. American Association for Public Opinion Research (AAPOR), Standard definitions: Final dispositions of case codes and outcome rates for surveys (4th ed.), 2006. Available at: http://www.aapor.org/pdfs/newstandarddefinitions.pdf.
27. Nagler RH, Gray, Romantan A, Kelly BJ, DeMichele A, Armstrong K, Schwartz JS, Hornik RC. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. *Patient Education and Counseling* 2010; **81**:S54–S62.
28. Dillman DA. *Mail and Internet Surveys: The Tailored Design Method*, 2nd Edition. John Wiley and Sons: New York, 2000.
29. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. J. Wiley and Sons: New York, 1987.
30. Kendall MG. *Rank Correlation Methods*, 4th Edition. Hodder Arnold, 1976.

