A Joint Modeling Approach for Longitudinal Outcomes and Non-ignorable Dropout under Population Heterogeneity in Mental Health Studies

Jung Yeon Park, PhD
Division of Educational Psychology and Research Methods
George Mason University
Address: West Building 2204,
4400 University Dr., Fairfax, VA 22030, United States
Email: jpark233@gmu.edu
Phone: 703-993-3691

Melanie M. Wall, PhD
Department of Biostatistics and Psychiatry
Columbia University
New York State Psychiatric Institute

Irini Moustaki, PhD
Department of Statistics
London School of Economics and Political Science

Arnold H. Grossman, PhD
Department of Applied Psychology
New York University

A Joint Modeling Approach for Longitudinal Outcomes and Non-ignorable Dropout under Population Heterogeneity in Mental Health Studies

Abstract

The paper proposes a joint mixture model to model non-ignorable drop-out in longitudinal cohort studies of mental health outcomes. The model combines a (non)-linear growth curve model for the time-dependent outcomes and a discrete-time survival model for the drop-out with random effects shared by the two sub-models. The mixture part of the model takes into account population heterogeneity by accounting for latent subgroups of the shared effects that may lead to different patterns for the growth and the drop-out tendency. A simulation study shows that the joint mixture model provides greater precision in estimating the average slope and covariance matrix of random effects. We illustrate its benefits with data from a longitudinal cohort study that characterizes depression symptoms over time yet is hindered by non-trivial participant drop-out.

Keywords: Latent growth curve, MNAR drop-out, survival analysis, finite mixture model, mental health.

1. Introduction

In many prospective cohort studies of mental health, it is often the case that some subject groups who have participated in the initial phase of the study fail to engage in a given sampling occasion, and subsequently never return on the next occasions. For example, when following a high-risk group for suicidal ideation (e.g. sexual minority youth followed across time), it is common that a particular subject group is less likely to stay engaged in the study. It can be hypothesized that the subjects with mental health problems are found to be more likely to drop out over time because of the severity of their mental illness condition or due to potentially changing risk characteristics making it difficult for them to return to the ongoing study. Increasing drop-out rates over time naturally generates missing data that may have an impact on statistical inference when analyzing dynamic changes in the outcome of interest (e.g. functioning or symptoms). Under missing completely at random (MCAR) and missing at random (MAR) missing data mechanisms (Little & Rubin, 2002; Rubin, 1976) where the process that causes the drop-out does not depend on either the observed or missing data (MCAR), or depends only on the observed data (MAR), approaches such as multiple imputation (MI), inverse probability weighting (IPW), or full information maximum likelihood (FIML) estimation using all available data can be used. But, if the drop-out mechanism depends largely on the unobserved or missing outcomes (e.g., on the values that are not collected due to the fact that the subject dropped out) even after controlling for all the observed variables, then the data are missing not at random (MNAR) and the drop out mechanism is considered to be “non-ignorable”. When drop-out is non-ignorable, the drop-out mechanism needs to be explicitly modeled together with the longitudinal outcomes and covariates.

Let the random variable Y_{it} denote the response for subject i at occasion t $(i = 1, \dots, N, t = 1, \dots, T)$. We also define a variable of missingness indicators, D_{it} such that $D_{it} = 0$ when Y_{it} is observed and $D_{it} = 1$ if a respondent drops out at time t. The vector Y_i is divided into observed $Y_{obs,i}$ and missing $Y_{mis,i}$ components respectively.

Under non-ignorable missingness, one needs to specify the joint density, $f(y_i, d_i)$. Three modeling approaches have been proposed to address non-ignorable drop-out. First, the *pattern-mixture* model (Wu & Carroll, 1988; Little, 1993; Hedeker & Gibbons, 1997; Demirtas & Schafer, 2003) divides the subjects into groups based on their drop-out patterns giving $f(y_i, d_i) = f(d_i)f(y_i|d_i)$. Second, the *selection* modeling approach (Diggle & Kenward, 1994) requires a model for the missing indicators giving $f(y_i, d_i) = f(y_i)f(d_i|y_i)$. Diggle and Kenward (1994) combined a multivariate linear model for the observed longitudinal responses with a logistic regression model for the non-ignorable drop-out process. Thirdly, the *shared-parameter* modeling approach (Roy, 2003) introduces a set of random effects (say b_i) that explain the interdependencies between the observed responses and the missing data indicator variables. In this case, $f(y_i, d_i) = \int_R f(y_i|b_i)f(d_i|b_i)\ db_i$.

The shared-parameter approach has received a lot of attention as it fits within the framework of latent growth curve modeling (LGC; Fitzmaurice, Laird, & Ware, 2004). More specifically, the same set of random effects (b_i) are used to capture individual differences in change over time (or “growth”), and to model the MNAR drop-out, i.e., $f(d_i|b_i)$. It is the shared random effects (that capture between-person differences in within-person change) between the two parts of the model that allow the joint model to handle MNAR. In our working example from a mental health study, this model is intuitive as we might expect the characteristics that drive change in depression symptoms over time to also affect drop-out. In this paper, we further extend the shared-parameter model to allow for latent subpopulations (i.e., mixtures or latent classes) to have their own developmental trajectory over time that may also be associated with the drop-out pattern. Mixtures have also received a lot of attention in the literature in order to address the problems caused by non-normality of the latent distribution

(Neuhaus, Hauck, & Kalbfleisch, 1992; Rizopoulos, Verbeke, & Molenberghs, 2008; Verbeke & Lesaffre 1996, 1997; Verbeke & Molenberghs, 2013; Baghfalaki, Ganjali, & Verbeke, 2017). For LGC analysis, Enders (2011) discussed the fact that even minor violations of population homogeneity can introduce substantial bias in key model parameters, and therefore a sensitivity analysis of the distributional assumptions (i.e., growth mixture) for the random intercept and slope is needed. Mixtures have also been considered within the context of non-ignorable missing data. Specifically, Muthen, Asparouhov, Hunter, and Leuchter (2011) extended the pattern-mixture and selection models by incorporating latent classes to identify different types of growth trajectories and demonstrated how this model framework could be used to check for MNAR vs MAR assumptions.

The present study examines the way in which the shared-parameter model with latent mixtures is able to estimate key model parameters in the presence of non-ignorable drop-out compared to a model that ignores population heterogeneity or ignores drop-out altogether. Specifically, we examine the shared parameter mixture model that integrates a discrete-time survival analysis model in the LGC model and demonstrate its applicability to handle MNAR drop-out. The rest of the paper is organized as follows. In Section 2, we describe the modified framework of an LGC model shared with a survival model for drop-out indicators, followed by a latent mixture model for the random intercept and slope. Section 3 demonstrates the performance of the model on the key parameters through three simulation scenarios. Then, Section 4 presents its applicability to a real-life data example modeling depression symptoms over time. Section 5 ends with our findings and concluding remarks.

2. The shared-parameter LGC and discrete time survival model for non-ignorable drop-out

2.1 Modeling the observed outcomes for a homogeneous population

The linear latent growth curve (LGC) model is written as:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad i = 1, \dots, N \qquad (1)$$

where Y_i is a $(T_i \times 1)$ vector of responses for the i^{th} subject and T_i is the number of observations for subject i. X_i is a $(T_i \times J)$ design matrix of the explanatory variables of the i^{th} subject and J is the number of explanatory variables including a column of 1s. More specifically, X_i contains the time scores and baseline explanatory variables (e.g. gender, socioeconomic status, age) written as $X_i = \left([\mathbf{1}], \left[X_{time,i,t}\right], \left[X_{base,i}\right]\right)$, where $[\mathbf{1}]$ denotes a column vector of 1s and $\boldsymbol{\beta} = \left(\beta_{10}, \dots, \beta_{1J}\right)^T$ is the vector of fixed effect coefficients. $Z_i = \left([\mathbf{1}], \left[X_{time,i,t}\right]\right)$ is a $(T_i \times 2)$ design matrix for a set of subject-specific random effects; and $\boldsymbol{b_i} = (b_{0i}, b_{1i})^T$ is a vector of random effects that represent a random intercept and random slope for each subject.

Similarly, the linear LGC given in (1) can be extended to allow for *non-linear* terms. For example, the non-linear growth model with quadratic terms is written as:

$$y_{it} = \beta_{10} + \beta_{11}\, x_{time,i,t} + \beta_{12}\, x^2_{time,i,t} + \sum_{j=3}^{J} \beta_{1j} x_{base,i,j} + b_{0i} + b_{1i}\, x_{time,i,t} + b_{2i}\, x^2_{time,i,t} + \varepsilon_{it}. \qquad (2)$$

The LGC model typically assumes population homogeneity of the random effects. Typically the $\boldsymbol{b}_i$ follows a multivariate normal distribution with mean $\boldsymbol{\mu_b} = 0$ and a variance-covariance matrix $\boldsymbol{G_b}$ i.e., $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{G_b})$, $\boldsymbol{\varepsilon}_i$ is a $(T_i \times 1)$ vector of residuals for the i^{th} subject, where $\boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \boldsymbol{V}_i)$ and $Cov(\boldsymbol{b}_i, \varepsilon_i) = 0$.

2.2 Modeling the Dropout

We define the probability that a respondent drops out at time t, given that they have remained in the study up to and including time $t-1$, by the hazard function $h_t = P(K = t | K \geq t), t = 2, \dots, T$ where K is a discrete random variable that indicates the time of the dropout. We have already defined the missing indicators, D_{it} such that $D_{it} = 0$ when Y_{it} is observed and $D_{it} = 1$ if a respondent drops out at time t (Moustaki & Steele, 2005; Muthén & Masyn, 2005). After the time of dropout, D_{it} itself is regarded as missing and can be set to an arbitrary value such as 999 or NA. We also consider the observations at the first occasion as complete data, so that $D_{i1} = 0$ and define $D = (D_{i2}, \dots, D_{iT})$. For example, in a study with four waves ($T = 4$), $D_i = \{0,0,1, NA\}$ if the subject dropped out at time $t = 3$. Similarly, $D_i = \{0,0,0,0\}$ if the subject never dropped out of the study. A subject who did not drop out by the end of the study will be the equivalent of censoring in survival analysis in which the event did not occur by the end of the study period.

With this notation, the hazard function can also be expressed as

$$h_{it} = P(K_i = t | K_i \geq t) = P(D_{it} = 1), t = 2, \dots, T. \tag{3}$$

Therefore, the probability of survival at each occasion t ($D_{it} = 0$) is the same as $(1 - h_{it})$. The model is then a sequence of these conditional probabilities, which makes the likelihood look identical to the situation where the observations at different times for the same person are independent.

The hazard function given in (3) is modeled as a function of covariates and the random effects $\boldsymbol{b}_i$ from (1) or (2) using any link function appropriate for binary data, including the logit and the probit functions. For example, if a quadratic model is fit (2) to the outcomes, the logit mixed-effect model for the dropout is written as:

$$logit\ h_{it}(\boldsymbol{x}_{it}, \boldsymbol{b}_i) = \ logit\{P(D_{it} = 1 | \ \boldsymbol{x}_{it}, \boldsymbol{b}_i)\} =$$

$$\beta_{20} + \beta_{21} x_{time,i,t} + \beta_{22} x_{time,i,t}^2 + \sum_{m=3}^{M} \beta_{2m} x_{base,i,m} + \gamma_0 b_{0i} + \gamma_1 b_{1i} + \gamma_2 b_{2i}, \tag{4}$$

where $\boldsymbol{\beta_2} = (\beta_{20}, \dots, \beta_{2M})^T$ is the vector of fixed effect coefficients, and $\boldsymbol{b}_i = (b_{0i}, b_{1i}, b_{2i})^T$ represents the random intercept and random slopes (linear and quadratic) for each subject as in (1) and (2). Notice that $\boldsymbol{b}_i$ is the vector of "shared" random effects, and $\boldsymbol{\gamma} = (\gamma_{0,}\gamma_1, \gamma_2)^T$ is the vector of regression coefficients that estimate the effects of $\boldsymbol{b}_i$ on the drop-out probability. If $\boldsymbol{\gamma}$ is zero, it implies that the probability of drop-out and the longitudinal responses are independent, and that the drop-out mechanism is ignorable. To the extent that at least one of the elements in γ is non-zero, the drop-out mechanism creates non-ignorable missing data for Y because it depends upon unobservable characteristics of Y summarized in the random effects vector $\boldsymbol{b}_i$.

2.3 Extending the shared parameter model with non-ignorable drop-out to include mixtures

It is well studied that when there is population heterogeneity in the subject-specific intercept or slopes in the LGC model, the standard assumption that $\boldsymbol{b}_i$ follows a (multivariate) normal distribution may lead to serious estimation biases in the fixed effect parameters and variances of the random effects (see Litiere, Alonso, & Molenberghs, 2008; Verbeke & Molenberghs, 2000; Verbeke & Lesaffre, 1996). Therefore, bias will also occur when making the homogeneity assumption in the proposed shared parameter model (e.g. (2) and (4)). A more flexible way of modeling the distribution of the random components is via a mixture model. Specifically, assume that the population consists of M subgroups of relative group sizes, ω_m, $\sum_{m=1}^{M} \omega_m = 1$, and in each group the subject-specific random effects (e.g., $b_{0i(m)}$,$b_{1i(m)}$, $b_{2i(m)}$) come from a multivariate normal distribution with different means, $\boldsymbol{\mu}_{\boldsymbol{b}_{(m)}}$ and a variance-covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{b}_{(m)}}$. More specifically, the distribution for the random effect vector, $\boldsymbol{b}_i = (b_{0i(m)}, b_{1i(m)}, b_{2i(m)})$ is approximated by a finite mixture of normal distributions given by:

$$\boldsymbol{b}_i \sim \sum_{m=1}^{M} \omega_m \, N\left(\boldsymbol{\mu}_{\boldsymbol{b}_{(m)}} = \begin{bmatrix} \mu_{b_0(m)} \\ \mu_{b_1(m)} \\ \mu_{b_2(m)} \end{bmatrix}, \; \boldsymbol{\Sigma}_{\boldsymbol{b}_{(.)}} = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0,b_1} & \sigma_{b_0,b_2} \\ \sigma_{b_0,b_1} & \sigma_{b_1}^2 & \sigma_{b_1,b_2} \\ \sigma_{b_0,b_2} & \sigma_{b_1,b_2} & \sigma_{b_2}^2 \end{bmatrix} \right), \qquad (5)$$

where $E(b_{0i}) = \sum_{m=1}^{M} \omega_m \, \mu_{b_0(m)}$ for the random intercept; $E(b_{1i}) = \sum_{m=1}^{M} \omega_m \, \mu_{b_1(m)}$ and $E(b_{2i}) = \sum_{m=1}^{M} \omega_m \, \mu_{b_2(m)}$ for the random slopes of the linear and the quadratic term respectively. The covariance matrices across the m subgroups are assumed equal to reduce the number of parameters and facilitate computation, hence the overall variance covariance matrix is given by:

$$\boldsymbol{G_b} = \sum_{m=1}^{M} \boldsymbol{\mu_{b(m)}} \boldsymbol{\mu_{b(m)}^T} \omega_m (1 - \omega_m) + \boldsymbol{\Sigma_b}. \qquad (6)$$

2.4 Estimation and Model Evaluation

Under the assumptions that the longitudinal outcomes Y_{it} and drop-out indicators D_{it} are conditionally independent given the random effects, the parameters associated with the observed responses and those with the drop out model are separated and the subjects are independent, the joint density is (shared-parameter):

$$f(y_i, d_i) = \int_R f(y_i|b_i) f(d_i|b_i)\ db_i \qquad (7).$$

The estimation of the proposed model was carried out by robust maximum likelihood estimation via expectation-maximization (EM) algorithm which can be implemented using Mplus Version 8.3 (Muthén & Muthén, 1998–2017); the code is available from https://github.com/pqcda/pqcda. More specifically, estimation of the proposed joint mixture model was carried out by full information maximum likelihood estimation via the EM algorithm; and robust standard errors were obtained. The convergence criteria for the EM algorithms include loglikelihood change (.0001), relative loglikelihood change (.0000001), and derivative (.0001) and the solution is given only if all convergence criteria are satisfied. For the approximation of the multidimensional integrals in (7) we used 15 integration points carried out with adaptive quadrature (Schilling & Bock, 2005).

Model selection criteria such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and Sample size Adjusted BIC (SABIC) that replaces sample size N by $(N+2)/24$, the BIC adjusted to prevent the BIC to underestimate the number of latent classes for small sample sizes (Nylund, Asparouhov, & Muthén, 2007), are used as relative fit indices where lower values indicate a better model. The model selection criteria will be used for selecting among different joint modeling approaches for MNAR. Because the standard LGC model under MAR does not directly model the drop out indicators, it effectively uses a different set of data (longitudinal outcomes without drop out indicators); therefore, likelihood values are not directly comparable to those from the joint models which hypothesize a specific form for the drop out indicators. Instead, the significance of the $\boldsymbol{\gamma}$ elements in the drop out model will primarily be examined to determine the presence of MNAR.

3. Simulated examples

3.1 Design

To study the performance of the proposed joint mixture model under the circumstance of non-ignorable drop-out, a series of simulated examples is conducted under three different scenarios. In all the cases, one simulated data set is analyzed. The number of occasions is taken to be four ($T = 4$) with sample size $N = 1{,}000$ subjects. To generate longitudinal outcomes in Scenarios 1 and 2, we use a linear growth curve model where the true values for the intercept and the slope parameters are taken to be $\beta_{10} = 30$ and $\beta_{11} = 1$, respectively. For Scenario 3, we use a quadratic growth curve model with parameters as in (2) with $\beta_{10} = 30$ and $\beta_{11} = 1$, and $\beta_{12} = 0.5$. To mimic situations where subjects' random effects (e.g., having higher outcome measures or experiencing faster change over time) influence the risk of dropping out, we use the logit mixed-effects model to generate the discrete-time drop out indicators (as in (4)) in which the probability of dropping out at each occasion increases as a function of the set of shared random effects, b_{0i}, b_{1i} (and b_{2i} for Scenario 3). In addition to the factor that examines (a) whether the dropout missingness of the survival process changes linearly or non-linearly across occasions, the three scenarios were also differentiated by two more factors: (b) underlying distributions assumed for the random effects; and (c) proportion of subjects falling in the latent classes (mixtures). Finally, the error term was randomly generated from a standard normal distribution in all scenarios.

Scenario 1. The population of subjects is heterogeneous at baseline only (b_{0i}). A mixture of bivariate normal distributions for the random intercept and slope (i.e., b_{0i} and b_{1i}) is used to characterize two latent subgroups with mean vectors $\boldsymbol{\mu} = (-10.5,0)^T$ and $\boldsymbol{\mu} = (3.5,0)^T$ for latent classes 1 and 2 respectively. The true proportions of the latent classes are $\omega_1 = 0.25$ and $\omega_1 = 0.75$, respectively. The true variance-covariance matrix within each class is $\boldsymbol{\Sigma_b} = \begin{bmatrix} 10 & -1.5 \\ -1.5 & 5 \end{bmatrix}$. The parameters associated with the drop out model are $\gamma_0 = 0.2$, and $\gamma_1 = 0.5$ (i.e., the probability of dropping out is greater for subjects with high levels of symptoms at baseline or as the symptoms get worse over time) and $\beta_{20} = -1$ and $\beta_{21} = 1$ (i.e., the risk of dropping out increases over time in the population). Figure 1 shows the distribution of Y (e.g. depression score) generated under Scenario 1 across the four waves. The dotted lines indicate the symptom score distribution if the drop-out missingness had not occurred. In contrast, the solid lines indicate the score distribution after the drop-out occurred. That is, difference between the two types of densities shows the effect of non-ignorable drop-out on the outcome variable. This Scenario 1 results in a drop-out rate of approximately 60% (across the four waves).

Scenario 2. The population is heterogeneous both at baseline (b_{0i}) and growth (b_{1i}). A mixture of bivariate normal distributions for the random intercept and slope (i.e., b_{0i} and b_{1i}) is used to characterize two latent subgroups with mean vectors $\mu_1 = (7.5, -0.5)^T$ and $\mu_2 = (-7.5, 0.5)^T$ for latent class 1 and 2 respectively. The latent class proportions are $\omega_1 = \omega_2 = 0.5$. The values chosen imply a heterogeneous population in which half of the subjects have high levels of symptoms that get better than average over time; and the other half have low levels of the symptoms that get worse over time. The true variance-covariance matrix in each class is $\boldsymbol{\Sigma_b} = \begin{bmatrix} 10 & -3 \\ -3 & 5 \end{bmatrix}$ and the true overall variance-covariance matrix using (6) becomes $\boldsymbol{G_b} = \begin{bmatrix} 38.13 & -4.88 \\ -4.88 & 5.12 \end{bmatrix}$. The data generation for drop-out occasions remain the same as in

Scenario 1 (i.e., $\gamma_0 = 0.2$, and $\gamma_1 = 0.5$ for the shared random effects and $\beta_{20} = -1$ and $\beta_{21} = 1$ for the fixed effects). Figure 2 shows bimodal densities of the depression scores with almost equal size groups in wave 1, turning into unequal size groups in wave 2, and a unimodal density in waves 3 and 4 due to the disappearance of one group with a greater average depression score. Besides, with the systematic drop-out, the score distribution in the presence of drop-out by wave 4 becomes much less variable than in the previous waves and compared to the true one. Scenario 2 results in approximately 54% missing data by wave 4.

Scenario 3. In this Scenario, we simulate the longitudinal outcomes from a quadratic LGC model such that the missing outcomes are interrelated with the longitudinal outcomes by means of the random slopes for the linear and quadratic terms. We also allow for the missing outcomes (drop-out) themselves to change non-linearly over time. Therefore, the dropout indicators are generated from model (4) that also includes a quadratic effect over time. The random intercept and the random slopes (i.e., b_{0i}, b_{1i}, and b_{2i}) are generated from a mixture of bivariate normal distributions with two latent subgroups. The mean vectors are $\mu_1 = (-10.5, 0.33, -0.1)^T$ and $\mu_2 = (3.5, -0.11, 0.03)^T$ for latent class 1 and 2 respectively, and latent class membership proportions of $\omega_1 = 0.25$ and $\omega_2 = 0.75$. The true variance-covariance matrix in each class is $\boldsymbol{\Sigma_b} = \begin{bmatrix} 10 & 3 & 1 \\ 3 & 3 & 0.5 \\ 1 & 0.5 & 1 \end{bmatrix}$ and true overall variance-covariance $\boldsymbol{G_b} = \begin{bmatrix} 32.97 & 2.28 & 1.22 \\ 2.28 & 3.02 & 0.49 \\ 1.22 & 0.49 & 1 \end{bmatrix}$. For the drop-out model (4) , the fixed intercept and slope are $\beta_{20} = -1$, $\beta_{21} = 2$, and $\beta_{22} = 0.5$, indicating that the risk of drop-out increases more drastically as a function of time compared to Scenarios 1 and 2. Furthermore, $\gamma_0 = 0.2$, $\gamma_1 = 1.5$, and $\gamma_2 = 1.5$ which implies that subjects with high levels of baseline symptoms ($0.2b_{0i}$) that grow faster over time ($1.5b_{1i} + 1.5b_{2i}$) tend to have high risks of dropping out of the study. As seen in Figure 3, with the systematic drop-out, the average score decreases over time (solid lines), but the variations remain the same, unlike Scenario 2. Scenario 3 results in approximately 43% with drop-out before wave 4.

3.2 Results

Tables 1-3 present the results from the three scenarios. Overall, in all three scenarios, the parameters of the fixed and random parts, as well as the class probabilities for the data-generating model (the 2-class joint mixture model), were well recovered. Furthermore, the data-generating model was consistently selected as the best-fitting model.

Table 1 presents the results of a single simulated data set from Scenario 1 where the random intercept (b_{0i}) is heterogeneous but the random slope (b_{1i}) is homogeneous. The parameter estimates of the shared random intercept (b_0) and slope (b_1) are statistically significant in both of the 1- and 2-class joint mixture models: $\hat{\gamma}_0 = 0.20, SE(\hat{\gamma}_0) = 0.01; \hat{\gamma}_1 = 0.47, SE(\hat{\gamma}_1) = 0.06$ and $\hat{\gamma}_0 = 0.25, SE(\hat{\gamma}_0) = 0.03; \hat{\gamma}_1 = 0.53, SE(\hat{\gamma}_1) = 0.07$ respectively. While the average intercept (β_{10}) was recovered by all models, the average slope (β_{11}) of the standard LGC model (under the MAR assumption) produces a noticeable large negative bias concluding $\hat{\beta}_{11} = -0.35$ and $SE(\hat{\beta}_{11})$=0.16 while the true value is $\beta_{11} = 1$. Also, it is found that both the model that ignores the non-ignorable drop-out mechanism (i.e. standard LGC) and the model that wrongly assumes homogeneity for the random intercept (1-class model), tend to result in considerable overestimation bias in the variance of the random intercept, $\hat{\sigma}_{b_0}^2 = 45.57$ and $SE(\hat{\sigma}_{b_0}^2) = 1.92$ whereas the true value is $\sigma_{b_0}^2 = 32.97$.

Table 2 presents the results from Scenario 2 that assumes that both the random intercept (b_{0i}) and random slope (b_{1i}) follow a mixture of two bivariate normals with different mean vectors in the two subpopulations. The standard LGC model (under MAR assumption) again underestimates, $\hat{\beta}_{11} = -0.23$ and $SE(\hat{\beta}_{11}) = 0.20$ (where $\beta_{11} = 1$) with a considerable bias. In contrast, both of the joint models which take into account drop-out (1- and 2-class models) give a negligible estimation error. However, the variance estimates of the random intercept

($\sigma_{b_0}^2$) from the incorrect models, tend again to result in considerable overestimation estimation: $\hat{\sigma}_{b_0}^2 = 66.06$ and $SE\left(\hat{\sigma}_{b_0}^2\right) = 1.82$ (standard LGC); and $\hat{\sigma}_{b_0}^2 = 66.65$ and $SE\left(\hat{\sigma}_{b_0}^2\right) = 1.72$ (joint model with 1-class) compared to the true value 38.13. The estimated link parameters (γ_0 and γ_1) of the two fitted joint models correctly show that the probability of dropping out on a given occasion increases for those subjects whose levels of the disorder are high to begin with or grow more rapidly over time. As seen in latent subgroups for the joint mixture (2-class) model, the estimated means for the latent classes correctly identified two trajectories: high levels of the disorder at baseline ($\mu_{b_0(1)}$) that tends to improve slowly relative to the average participant over time ($\mu_{b_1(1)}$); or low levels of the disorder at baseline ($\mu_{b_0(2)}$) that grow more rapidly than the average participant's over time ($\mu_{b_1(2)}$).

Finally, Table 3 gives the results from Scenario 3 that assumes that the random intercept (b_{0i}), linear slope (b_{1i}), and quadratic slope (b_{2i}) together follow a mixture of two bivariate normals with different mean vectors in the two subpopulations. In this scenario, we fit linear and quadratic joint models (1-class) and (2-class). Under the standard LGC model (under MAR assumption) we find considerable negative bias for the fixed slope, $\hat{\beta}_{11} = -1.25$ and $SE\left(\hat{\beta}_{11}\right) = 0.12$ that fails to cover the true value (β_{11}= 1) within the 95% CI. The misfit of the model is also revealed by the significance of the coefficients (γ) of the shared random effects in the joint model. We find again that the estimated variance of the random intercept in the standard model and the joint model with 1-class largely overestimate the true value (true overall variance is $\sigma_{b_0}^2$=32, i.e., $\sigma_{b_0}^2$=10 per class). The joint mixture model with 2-classes correctly estimates means for two trajectories: lower levels of the disorder at baseline ($\mu_{b_0(1)}$) that leave room for it growing more rapidly than the average participant's over time ($\mu_{b_1(1)}$); or high levels of the disorder at baseline ($\mu_{b_0(2)}$) that grow more slowly than the average participant's over time ($\mu_{b_1(2)}$).

4. The shared-parameter LGC mixture model for depression symptoms

Our real-data example comes from a four-year longitudinal study designed to examine the risk and predictive factors for suicide via a yearly survey among sexual minority youth (15-21 age range) collected between 2011 and 2015. The participants were recruited in three U.S. cities (located in the Northeast, Southwest, and on the West Coast) from community organizations and college groups for LGBTQ+ youth. Among these three sites, the recruitment procedures in the West Coast region were applied ineffectively during the process of outreach and therefore, data are unavailable after the first wave. The sample sizes were n = 989 (in wave 1), 544 (in wave 2), 368 (in wave 3), and 305 (in wave 4), a 69.8% drop out by wave 4. Depression symptoms were measured with the Beck Depression Inventory–Youth (BDIY; Beck, Beck, Jolly, & Steer, 2005) which is a composite of 20 questions such as, “I think that bad things happen because of me” and “I think I do things badly”. Responses to those questions were scored as 0, 1, 2, and 3, indicating “never”, “sometimes”, “often”, and “always”, respectively and combined for a range of 0 to 100, with the typical cutoff suggested of greater than or equal to 55 which is considered to represent elevated depression. The survey also collected information on whether the respondents received free lunch at school (yes/no), a proxy measure for socioeconomic status. The average scores at wave 1 for those who dropped out after waves 1, 2, and 3 are 17.80, 17.30, and 17.17, respectively. The average score at wave 1 for those who never dropped out is 14.13, which is noticeably lower, indicating weaker depressive symptoms. We fit a total of five models examined in the simulated Scenario 3, including covariates. Table 4 shows model fit indices, parameter estimates, and standard errors for a standard LGC model (under MAR assumption) and two joint models (under MNAR or non-ignorable drop-outs assumption) with the linear and/or quadratic terms on the sub-models and with 1- and 2-classes. Among the MNAR models, the joint model with 2-class mixtures including only linear terms on the sub-models was selected as the best model (AIC: 18,360, BIC: 18,466, and ABIC: 18,396). The joint mixture model with quadratic terms did not improve the model fit and therefore it is not presented here. We discuss below the results from the selected model.

The joint mixture model (2-class) estimates 20% of the subjects (ω_1) as being classified in the first subgroup and 80% (ω_2) in the second subgroup. Figure 4 shows the depression scores identified by the two latent groups. The results suggest that in group 1 (solid red), subjects with a higher baseline depressive symptom tend to experience a more rapid decrease in the depressive symptom than an average subject across waves. Whereas, subjects in group 2 (solid grey) already begin with a lower baseline depression at the beginning, and therefore, tend not to show much of an improvement.

The results under “MNAR dropout” in Table 4 also show that the drop-out mechanism is non-ignorable. Specifically, for the joint mixture model (2-class), the estimated link parameter $\hat{\gamma}_1$ for the shared random slope was found to be significant in predicting the log hazard odds for drop-out ($\hat{\gamma}_1$= -0.08, *SE* ($\hat{\gamma}_1$)=0.03), suggesting that subjects who experienced a faster improvement in their depression score were more likely to drop out during the study. However, the effect of $\hat{\gamma}_0$ was not significant when $\hat{\gamma}_1$ is present.

Results of the model under the “latent growth process” section of Table 4 suggest that Northeast (“Site1”) has lower depressive symptoms as compared to the West Coast. SES was not found to be statistically significant in any of the three models. The estimated average slope ($\hat{\beta}_{11}$) is statistically significant and negative, indicating improvement (i.e. lessening) of depressive symptoms over time (e.g., $\hat{\beta}_{11}$=-0.95, *SE* ($\hat{\beta}_{11}$)=0.27). Similar to Simulation 3, we notice that the standard LGC model (MAR assumed) that does not model for potential MNAR dropout missingness leads to a considerable downward bias in the fixed slope ($\hat{\beta}_{11}$=-5.56, *SE* ($\hat{\beta}_{11}$)=0.20).

5. Discussion

The main contribution of the paper centers on extending the latent growth model to address the non-ignorable dropout missingness in multi-wave survey data where population heterogeneity is present. In particular, we are interested in identifying a particular type of population heterogeneity of the participants in order to answer the question: is there a non-ignorable dropout pattern for those who share a specific growth pattern in the longitudinal analysis? Therefore, we have presented here a modified shared-parameter (or joint) model in order to explain the interdependence inherent in a longitudinal response process and the risk of non-ignorable drop-out in the context of multi-wave mental health cohort studies. To have a full representation of the joint distribution, the model is comprised of two sub-models, a LGC model and a discrete-time survival model that are linked by a common set of random effects. The LGC model in the study has been generalized to allow for both linear and quadratic random slopes. Similarly, the logistic survival model with non-linear terms has been used to handle an additional complexity, namely that the drop-out patterns themselves are non-linear over time. In that framework, mixtures were used to approximate the distribution of the random effects and to identify population heterogeneity in the analysis.

Three simulated examples showed that the estimated values can be far from the truth when it comes to the key parameters (fixed slope in particular) when non-ignorable drop-out was not adequately modeled in the LGC analysis. Even with the shared-parameter model, however, we found that estimation error in the variance covariance matrix of the random effects persisted when normality was assumed in a heterogeneous population. To remedy this, the joint mixture model was able to remove the error due to the drop-out by weakening the normality assumption of the random effects that helped to identify latent subgroups based on the subject-specific trait of the mental disorder. Finally, we have focused on its practical application in a real-life data analysis in relation to depressive symptoms.

Our approach is implemented using the software Mplus with a full information maximum likelihood estimation via the EM algorithm. With respect to determining the number of latent classes, we have found that AIC, BIC, and SABIC were able to choose the correct mixture model when there is a population heterogeneity and to test redundancy of non-linear terms of the modeling approach. We acknowledge that different model selection criteria may be worth exploring to be able to determine the optimal number of latent classes in the context of mixture modeling. For example, Gelman, Hwang, and Vehtari (2014) discussed that Watanabe-Akaike information criteria (WAIC; Watanabe, 2010) is viewed as a more stable information criterion for selecting optimal hierarchical and mixture structures, utilizing the entire posterior distribution, rather than point estimates, in estimating the penalty. Also, based on a posterior predictive distribution to ascertain how many latent classes are optimal for predicting drop-out occasions for individuals, Mean Square Predictive Error (MSPE) can be used to predict drop-out occasions to choose between competing models. It may easily be possible to implement the present model in such a program as Stan (Stan Development Team, 2018), but doing so was beyond the scope of the present manuscript that meant to examine the model itself, not to provide all possible ways of performing estimation. As extended studies, it will be interesting to use the proposed model to nonlinear growth curves for multivariate longitudinal outcomes when the MNAR is present in the mental health study. Furthermore, the random effects/latent variables can be multidimensional in this context, for example, Hafez, Moustaki, and Kuha (2015) used a continuous time-dependent latent variable to explain the associations among multiple categorical outcome variables and a separate continuous latent variable to account for the interdependence among the drop-out indicators. Nevertheless, we believe that the current study provides researchers and practitioners with a reliable and flexible methodological tool for modeling longitudinal mental health and other data under non-ignorable drop-out.

Acknowledgements

This article uses data from the Risk and Protective Factors for Suicide among Sexual Minority Youth study designed by Arnold H. Grossman and Stephen T. Russell and supported by Award Number R01MH091212 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health. The authors express their gratitude to the study's participants for sharing their experiences.

References

1. Baghfalaki, T., Ganjali, M., & Verbeke, G. (2017). A shared parameter model of longitudinal measurements and survival time with heterogeneous random-effects distribution, *Journal of Applied Statistics, 44*:15, 2813-2836.
2. Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757-786.
3. Bauer, D. J., & Curran, P. (2003). Distributional assumptions of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods, 8*, 338-363.
4. Bauer, D. J., & Curran, P. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3-29.
5. Beck, J.S., Beck, A.T., Jolly, J.B., & Steer, R.A. (2005). Beck Youth Inventories-Second Edition for Children and Adolescents manual. San Antonio, TX: PsychCorp.
6. Beunckens, C., Molenberghs, G., Verbeke, G. & Mallincrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics, 64*, 96–105.
7. De Gruttola, V. & Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics, 50*, 1003–1014.

8. Diggle, P.D. & Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics, 43*:49–73.
9. Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*(1), 1-16.
10. Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2004). Applied longitudinal analysis. John Wiley & Sons, Hoboken.
11. Gelfand, A. & Ghosh, S. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika, 85*, 1–11.
12. Gelman, A. & Rubin, D. B. (1995). Avoiding model selection in Bayesian social research. *Sociological methodology, 25*, 165-173.
13. Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6*, 733-807.
14. Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*, 997-1016.
15. Glas, C. A. W. & Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educ. Psychol. Measmnt*, *68*, 907–922.
16. Hafez, M. S., Moustaki, I., & Kuha, J. (2015). Analysis of multivariate longitudinal data subject to nonrandom dropout. *Struct. Equn Modlng*, *22*, 193–201.
17. Knott, M., Albanese, M. T., & Galbraith, J. (1990). Scoring attitudes to abortion. *Statistician, 40*, 217–223.
18. Kuha, J., Katsikatsou, M., & Moustaki, I. (2018). Latent variable modelling with non-ignorable item non-response: multigroup response propensity models for cross-national analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* doi:10.1111/rssa.12350.

19. Laud, P. W. & Ibrahim, J. G. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society, Ser. B, 57,* 247-262.
20. Lesaffre, E. & Lawson, A. (2012). *Bayesian Biostatistics*, John Wiley & Sons; New York; 2012-01.
21. Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Ass.*, 90, 1112–1121.
22. Manwani, S.G., Szilagyi, K.A., Zablotsky, B., Hennen, J., Griffin, M.L., & Weiss, R.D. (2007). Adherence to pharmacotherapy in bipolar disorder patients with and without co-occurring substance use disorders. *J. Clin. Psychiatry, 68*, 1172-1176.
23. Moustaki, I. & Steele, F. (2005). Latent Variable Models for Mixed Categorical and Survival Responses, with an Application to Fertility Preferences and Family Planning in Bangladesh. *Statistical Modelling, 5,* 327-342.
24. Moustaki, I. & Knott, M. (2000). Weighting for Item Non-Response in Attitude Scales Using Latent Variable Models with Covariates. *Journal of the Royal Statistical Society, Series A, Vol. 163*(3), 445-459.
25. Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. Journal of *Educational and Behavioral Statistics, 30*, 27–58.
26. Muthén, L.K. & Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
27. Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16*(1), 17-33.
28. Nelson, F. D. (1977). Censored regression models with unobserved, stochastic censoring thresholds. *J. Econmetr., 6*, 309–327.

29. Neuhaus, J. M., Hauck, W. W., & Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models, Biometrika, 79, 4, 755–762.
30. Nylund, K.L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535-569.
31. Nossel, I., Wall, M.M., Scodes, J., Marino, L., Zikha, S., Bello, I., Malinovsky, I., Lee, R., Radigan, M., Smith, T., Sederer, L., Gu, G., & Dixon, L. (2018). Results of a coordinated specialty care program for early psychosis and predictors of outcomes. Psychiatric Services, 69(8): 863-870.
32. O'Muircheartaigh, O. & Moustaki, I. (1999). Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A., Vol. 162*, 177-194.
33. Pohl, S., Grafe, L. & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests: evaluating approaches accounting for missing responses in item response theory models. *Educ. Psychol. Measmnt*, *74*, 423–452.
34. Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). Latent growth curve modeling. Thousand Oaks, CA: Sage Publications.
35. Prisciandaro, J.J., Rembold, J., Brown, D.G., Brady, K.T., & Tolliver, B.K. (2011). Predictors of clinical trial dropout in individuals with co-occurring bipolar disorder and alcohol dependence, *Drug and Alcohol Dependence, 118*, 493-496.
36. Rizopoulos D., Verbeke G., & Molenberghs, G. (2008). Shared parameter models under random-effects misspecication, *Biometrika, 95*, 63-74.
37. Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics, 59*, 829–836.

38. Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*(3), 533-555.
39. Schluchter, M. (1992). Methods for the analysis of informatively censored longitudinal data. *Statist. Med.*, *11*, 1861–1870.
40. Stan Development Team. 2018. Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0. http://mc-stan.org
41. Su, Y. S., & Yajima, M. (2015). R2jags: Using R to run ‘JAGS’. R package version 0.5-7.
42. Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., & Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, 54, 367–383.
43. Verbeke, G., & Molenberghs, G. (2013). The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics*, *14*(3), 477-490.
44. Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Appl. Psychol. Measmnt*, *39*, 583–597.
45. Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
46. Wu, M. C. & Carroll, R. J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, *44*, 175–188.

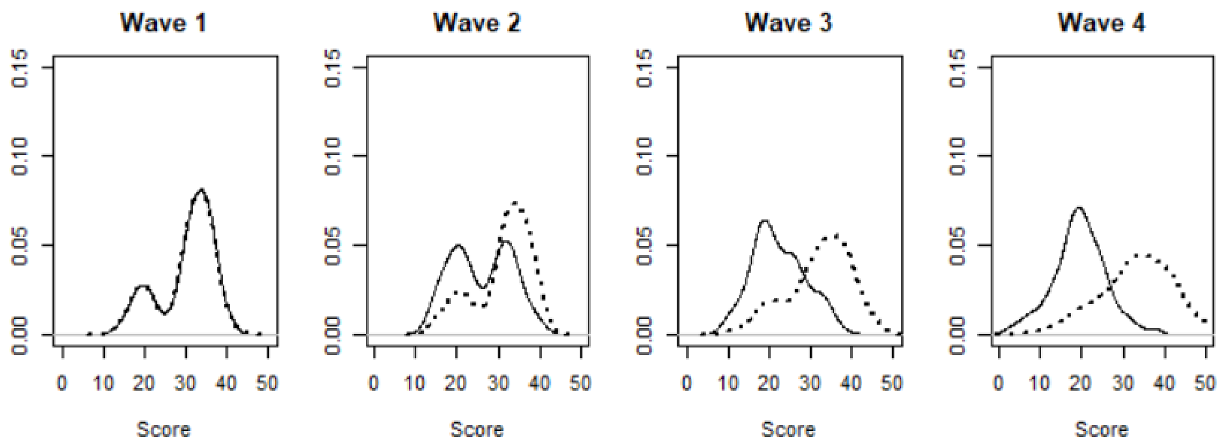

Figure 1. ***Scenario 1****: Distribution of symptom scores across four waves. The dotted lines indicate the symptom score distribution if the drop-out missingness had not occurred. In contrast, the solid lines indicate the score distribution after the drop-out occurred.*

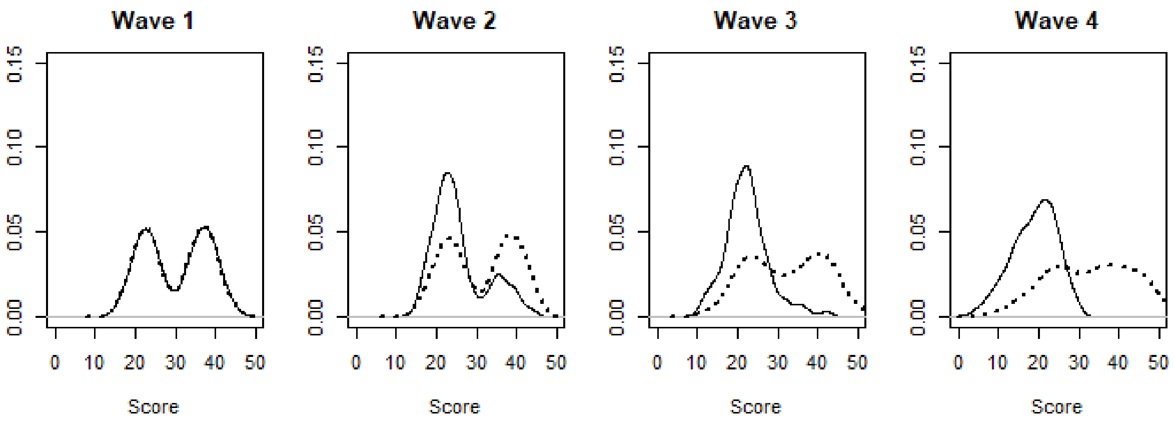

Figure 2. ***Scenario 2:*** *Distribution of symptom scores across four waves. The dotted lines indicate the symptom score distribution if the drop-out missingness had not occurred. In contrast, the solid lines indicate the score distribution after the drop-out occurred.*

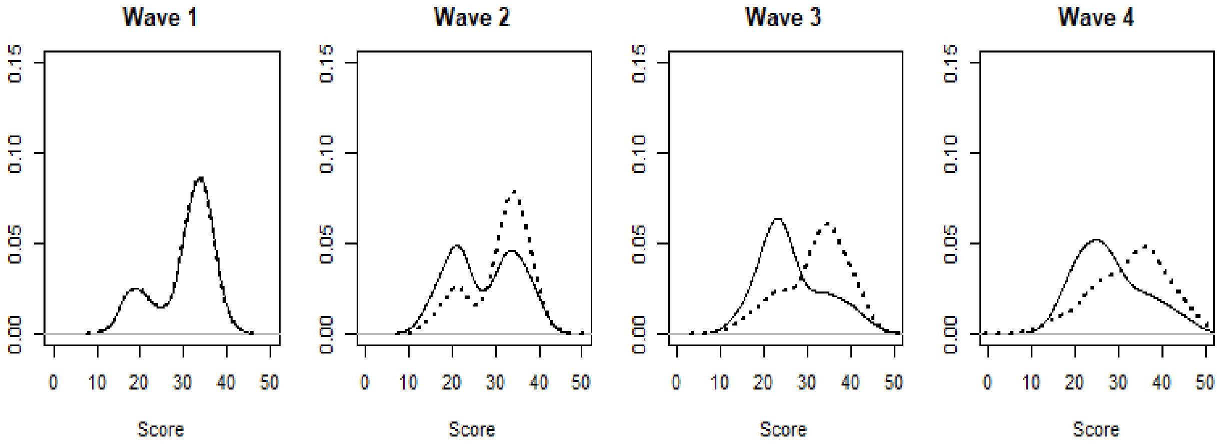

Figure 3. ***Scenario 3****: Distribution of symptom scores across four waves. The dotted lines indicate the symptom score distribution if the drop-out missingness had not occurred. In contrast, the solid lines indicate the score distribution after the drop-out occurred.*

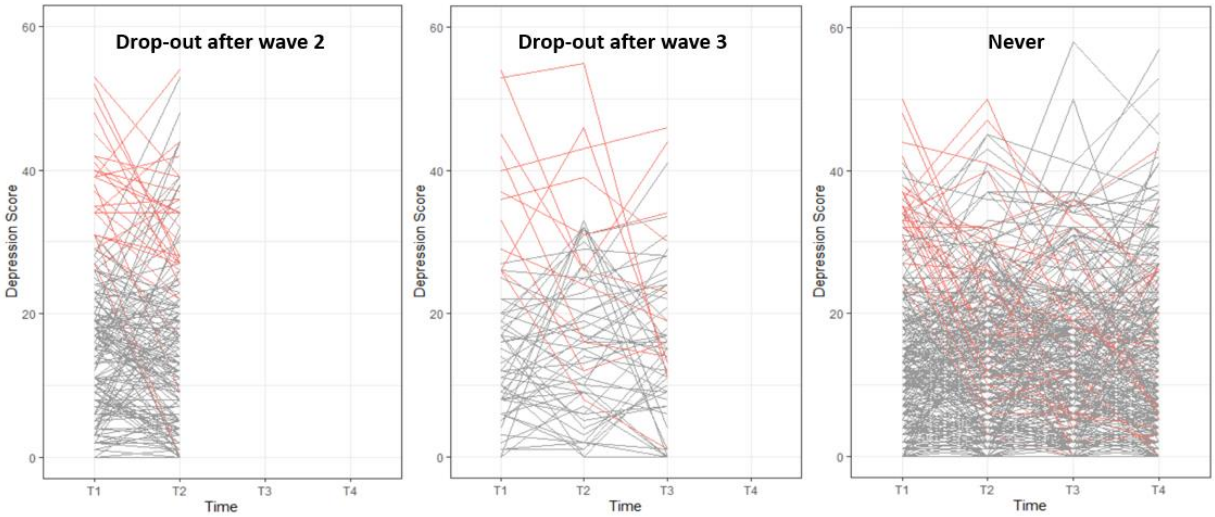

Figure 4. ***Depression data example****: Trajectories of depressive symptoms over time for subjects who dropped out after wave 2, wave 3, and never dropped (from far left) and latent class memberships estimated by joint models (2-class). Note. solid lines= results of the joint mixture model (2-class), where red and grey indicate two latent memberships.*

***Table 1**. Scenario 1: True and estimated parameter values, data generated from a joint model (2-class) with a bivariate normal for the heterogeneous random intercept.*

Parameter	True	Standard LGC (MAR assumed)		Joint model (1-class)		Joint model (2-class)	
		Est.	S.E.	Est.	S.E.	Est.	S.E.
Latent growth process							
β_{10}	30	30.18	0.22	30.17	0.22	30.17	0.22
$\boldsymbol{\beta_{11}}$	**1**	**-0.35**	**0.16**	**1.21**	**0.25**	**1.21**	**0.26**
$\boldsymbol{\sigma_{b_0}^2}$	**32.97**	**45.32**	**1.92**	**45.57**	**1.92**	**32.23**	**1.03**
$\sigma_{b_1}^2$	5	4.80	0.51	5.65	0.67	5.36	0.61
σ_{b_0,b_1}	-1.5	-2.97	0.82	0.38	0.96	-0.69	0.73
MNAR dropout							
γ_0	0.2	.	.	0.20	0.01	0.25	0.03
γ_1	0.5	.	.	0.47	0.06	0.53	0.07
Latent subgroups							
$\mu_{b_0(1)}$	-10.5	.	.	.	.	-10.72	0.22
$\mu_{b_0(2)}$	3.5	.	.	.	.	3.25	0.26
$\mu_{b_1(1)}$	0	.	.	.	.	-0.38	0.24
$\mu_{b_1(2)}$	0	.	.	.	.	0.12	0.07
$\sigma_{b_0(.)}^2$	10	.	.	.	.	9.79	0.57
$\sigma_{b_1(.)}^2$	5	.	.	.	.	5.38	0.61
$\sigma_{b_0,b_1(.)}$	-1.5	.	.	.	.	-1.41	0.49
ω_1	0.25	.	.			0.23	0.01
LL		-4785.57		-5566.19		-5369.82	
AIC		9589.15		11158.38		**10775.64**	
BIC		9633.32		11222.18		**10863.98**	
SABIC		9604.74		11180.90		**10806.81**	

Note. LGC=Latent Growth Curve; AIC=Akaike information criterion; BIC=Bayesian information criterion; SABIC=Sample-Size Adjusted Bayesian information criterion.

Table 2*. Scenario 2: True and estimated parameter values, data generated from a joint model (2-class) with a bivariate normal for the heterogeneous random intercept and slope.*

Parameter	True	Standard LGC (MAR assumed)		Joint model (1-class)		Joint model (2-class)	
		Est.	S.E.	Est.	S.E.	Est.	S.E.
Latent growth process							
β_{10}	30	29.79	0.26	29.77	0.26	29.77	0.26
$\boldsymbol{\beta_{11}}$	**1**	**-0.23**	**0.20**	**1.02**	**0.17**	**1.16**	**0.18**
$\boldsymbol{\sigma^2_{b_0}}$	**38.13**	**66.06**	**1.82**	**66.65**	**1.72**	**38.06**	**0.96**
$\sigma^2_{b_1}$	5.12	4.48	0.58	4.93	0.47	4.92	0.43
σ_{b_0,b_1}	-4.88	-7.73	1.53	-7.93	1.10	-5.17	0.70
MNAR dropout							
γ_0	0.2	.	.	0.21	0.02	0.21	0.03
γ_1	0.5	.	.	0.53	0.06	0.57	0.07
Latent subgroups							
$\mu_{b_0(1)}$	7.5	.	.	.	.	7.45	0.26
$\mu_{b_0(2)}$	-7.5	.	.	.	.	-7.72	0.27
$\mu_{b_1(1)}$	-0.5	.	.	.	.	-0.64	0.14
$\mu_{b_1(2)}$	0.5	.	.	.	.	0.66	0.14
$\sigma^2_{b_0(\cdot)}$	10	.	.	.	.	9.28	0.61
$\sigma^2_{b_1(\cdot)}$	5	.	.	.	.	4.71	0.44
$\sigma_{b_0,b_1(\cdot)}$	-3	.	.	.	.	-2.71	0.47
ω_1	0.5	.	.			0.51	0.02
LL		-4947.41		-6323.40		-6044.49	
AIC		9912.82		12672.80		**12124.97**	
BIC		9956.99		12736.60		**12213.31**	
SABIC		9928.41		12695.31		**12156.14**	

Note. LGC=Latent Growth Curve; AIC=Akaike information criterion; BIC=Bayesian information criterion; SABIC=Sample-Size Adjusted Bayesian information criterion.

***Table 3**. Scenario 3: True and estimated parameter values, data generated from a joint mixture model (2-class) with a bivariate normal for the heterogeneous random intercept and slopes for linear and quadratic terms.*

		Standard LGC (MAR assumed)		Joint model (1-class)				Joint model (2-class)			
		linear		linear		linear + quadratic		linear		linear + quadratic	
Parameter	True	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Latent growth process											
β_{10}	30	30.24	0.22	30.25	0.22	30.26	0.22	30.25	0.22	30.25	0.22
$\boldsymbol{\beta_{11}}$	**1**	**-1.25**	**0.12**	**1.83**	**0.37**	**0.98**	**0.29**	**1.88**	**0.19**	**0.80**	**0.28**
β_{12}	0.5	.	.	.	.	0.52	0.18	.	.	0.71	0.18
$\boldsymbol{\sigma^2_{b_0}}$	**32**	**53.44**	**2.59**	**46.42**	**2.38**	**43.65**	**2.06**	**31.64**	**10.96**	**31.58**	**1.00**
$\sigma^2_{b_1}$	3.02	9.67	0.86	8.73	2.06	3.12	0.80	8.68	0.98	2.50	0.66
$\sigma^2_{b_2}$	1	.	.	.	.	0.84	0.16	.	.	1.00	0.22
σ_{b_0,b_1}	2.28	-11.55	1.62	4.03	2.11	2.62	1.24	4.56	0.70	2.43	0.74
σ_{b_0,b_2}	1.22	.	.	.	.	1.09	0.55	.	.	1.50	0.45
σ_{b_1,b_2}	0.49	.	.	.	.	0.70	0.20	.	.	0.79	0.19
MNAR dropout											
γ_0	0.2	.	.	0.17	0.02	0.17	0.03	0.28	0.04	0.22	0.06
γ_1	1.5	.	.	0.87	0.09	1.26	0.20	0.96	0.09	1.22	0.39
γ_2	1.5	.	.	.	.	1.26	0.20	.	.	1.74	0.33
Latent subgroups											
$\mu_{b_0(1)}$	-10.5	.	.	.	.			-10.63	0.27	-10.64	0.28
$\mu_{b_0(2)}$	3.5	.	.	.	.			3.34	0.20	3.34	0.06
$\mu_{b_1(1)}$	0.33	.	.	.	.			0.10	0.21	0.13	0.20
$\mu_{b_1(2)}$	-0.11	.	.	.	.			-0.14	0.06	-0.04	0.06
$\mu_{b_2(1)}$	-0.1	.	.	.	.			.	.	-0.16	0.19
$\mu_{b_2(2)}$	0.03	.	.	.	.			.	.	0.05	0.03
$\sigma^2_{b_0}$	10	.	.	.	.	.	.	9.05	0.49	8.96	0.52
$\sigma^2_{b_1}$	3	.	.	.	.	.	.	8.68	0.98	2.50	0.67
$\sigma^2_{b_2}$	1							.	.	1.00	0.22
σ_{b_0,b_1}	3	.	.	.	.	.	.	4.36	0.49	2.70	0.50
σ_{b_0,b_2}	1	.	.	.	.	.	.	.	.	1.15	0.31
σ_{b_1,b_2}	0.5	.	.	.	.	.	.	.	.	0.79	0.19
ω_1	0.25	.	.					0.24	0.01	0.24	0.01
LL		-5845.34		-6768.82		-6890.48		-6603.97		-6479.99	
AIC		11708.69		13563.64		13816.97		13243.95		**13011.99**	
BIC		11752.86		13627.44		13905.31		13332.29		**13139.59**	
SABIC		11724.27		13586.15		13848.14		13275.12		**13057.02**	

Note. LGC=Latent Growth Curve; AIC=Akaike information criterion; BIC=Bayesian information criterion; SABIC=Sample-Size Adjusted Bayesian information criterion.

***Table 4**. Results of standard latent growth curve model (under MAR) and joint (mixture) models (under MNAR), Depression data example.*

	Standard LGC (MAR assumed)		Joint model (1-class)				Joint model (2-class)	
	linear		linear		linear + quadratic		linear	
Parameter	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Latent growth process								
Site 1	-5.43	0.83	-5.33	0.83	-5.33	0.83	-4.08	0.76
Site 2	-0.82	1.1	-0.6	1.11	-0.59	1.11	0.21	1.07
SES	0.27	0.7	0.16	0.7	0.16	0.7	-0.28	0.65
β_{10}	19.09	0.76	19.11	0.76	19.13	0.76	18.68	0.74
β_{11}	**-5.56**	**0.20**	**-0.59**	**0.22**	**-0.78**	**0.53**	**-0.95**	**0.27**
β_{12}	.	.	.	.	0.07	0.17	.	.
$\sigma^2_{b_0}$	**95.66**	**8.86**	**96.89**	**8.94**	**104.13**	**19.28**	**88.63**	**11.72**
$\sigma^2_{b_1}$	5.95	2.09	6.35	2.12	15.27	20.11	7.08	2.64
$\sigma^2_{b_2}$			.	.	0.35	0.78	.	.
σ_{b_0,b_1}	-8.04	4.11	-8.72	4.18	-18.44	20.83	-9.48	5.56
σ_{b_0,b_2}	.	.	.	.	2.64	5.09	.	.
σ_{b_1,b_2}	.	.	.	.	-2.00	4.30	.	.
MNAR dropout								
γ_0	.	.	0.01	0.01	0.01	0.01	0.003	0.01
γ_1	.	.	-0.03	0.03	-0.92	0.03	-0.08	0.03
γ_2	.	.	.	.	-0.004	0.06	.	.
Latent subgroups								
$\mu_{b_0(1)}$	.	.	.	.			-17.02	1.13
$\mu_{b_0(2)}$	.	.	.	.			4.25	0.67
$\mu_{b_1(1)}$	.	.	.	.			5.37	1.68
$\mu_{b_1(2)}$	.	.	.	.			-1.34	0.24
$\mu_{b_2(1)}$	.	.	.	.			.	.
$\mu_{b_2(2)}$	.	.	.	.			.	.
$\sigma^2_{b_0}$	.	.	.	.	.		39.55	11.32
$\sigma^2_{b_1}$			.	.			2.066	1.68
$\sigma^2_{b_2}$			.	.				
σ_{b_0,b_1}			.	.			6.17	2.91
σ_{b_0,b_2}			.	.				
σ_{b_1,b_2}			.	.				
ω_1			.	.			0.20	0.04
LL			-9208.08		-9202.59		**-9158.85**	
AIC	16371		18454.16		18455.19		**18359.71**	
BIC	16429		18547.20		18577.61		**18466.23**	
SABIC	16391		18486.85		18498.21		**18396.01**	

Note. LGC=Latent Growth Curve; Site 1=Northeast; Site 2=Southwest; SES=Socioeconomic Status (Free Lunch); AIC=Akaike information criterion; BIC=Bayesian information criterion; SABIC=Sample-Size Adjusted Bayesian information criterion.