Abstract

Regression mixture models are one increasingly utilized approach for developing theories about and exploring heterogeneity of effects. This study aims to extend the current use of regression mixtures to repeated regression mixture method when repeated measures, such as diary type and experience sampling method data, are available. We hypothesized that additional information borrowed from the repeated measures would improve the model performance in terms of class enumeration and accuracy of parameter estimates. We specifically compared three types of model specifications in regression mixtures: (a) traditional single outcome model, (b) repeated measures model with 3, 5, and 7 measures, and (c) single outcome model with the average of 7 repeated measures. Results showed that the repeated measures regression mixture models substantially outperformed the traditional and average single outcome model in class enumeration with less bias in parameter estimates. For sample size, whereas prior recommendations have suggested the regression mixtures require samples of well over 1000 participants even for the large distance classes (regression weight of .20 vs .70 classes), the current repeated measures regression mixture models allow for samples as low as 200 participants with increased number of (i.e., seven) repeated measures. Application of the proposed repeated measures approach is also demonstrated using the Sleep Research Project data. Implications and limitations of the study are discussed.

Keywords: Regression mixture models, sample size, repeated measures, heterogeneous effects

Repeated Measures Regression Mixture Models

Introduction

Regression mixture models are a novel approach to identify heterogeneous effects. These models detect two or more sub-populations, which differ in the effect of a predictor on an outcome, using a latent class variable instead of the *a priori* moderator(s). For example, regression mixture models were used to find a differential susceptibility to the effects of a child's temperament on the mother-child interaction style (Lee, 2013), where they found three latent classes with varying associations between the child's temperament and mother-child interactions (i.e., non-susceptible class, susceptible-high class, and susceptible-low class). In another study, regression mixture approach was used to understand the differential effects of depressive symptoms and neuroticism on cognitive complaints in older adults (Kliegel & Zimprich, 2005). Given the exploratory nature of this approach, regression mixture models have become increasingly popular in the last decade (Dyer, Pleck, & McBride, 2012; Lanza, Cooper, & Bray, 2014; Lanza, Kugler, & Mathur, 2011; McDonald et al., 2016; Schmiege & Bryan, 2016; Silinskas et al., 2013; Silinskas et al., 2016).

The current study introduces the use of regression mixture models with repeated measures. One of the commonly used mixture approaches to repeated measures is a growth mixture model (Muthén, 2004; Muthén & Muthén, 2000). Growth mixture models have been increasingly popular and applied in a wide range of fields including health, educational, and psychological studies. The main purpose of growth mixture models is to identify heterogeneity in growth trajectories assuming that the data are from a mixture of subpopulations of individuals with different growth trajectories. Growth mixture models identify these subpopulations by estimating class-specific growth parameters (e.g., intercept, change of rate, variance). Under the framework of multilevel modeling (i.e., repeated measures are nested within individuals), random-effects mixture models (Ng et al, 2006; Xu & Hedeker, 2001) have been also widely employed to unfold the possible subpopulation heterogeneity. Similar to the growth mixture models, the main interest in the use of random-effect mixture models has been typically on determining the heterogeneity in the growth trajectory of repeated outcomes over time (Kohli, Harring, & Zopluoglu, 2016; Yau, Lee, & Ng, 2003) in which the latent classes are identified based on the fixed and random effects on each time point.

While finite mixtures have been used extensively for examining heterogeneity in longitudinal trajectories (Bauer & Curran, 2003; Jung & Wickrama, 2008; Nylund, Asparouhov, & Muthén, 2007), our purpose in using repeated measures regression mixtures is quite different. Rather than focusing on trends over time, repeated measures are collected to increase the amount of information available about the outcome. Obvious uses of this approach are situations when repeated measures data are available but longitudinal trends are not of interest but considered as nuisance parameters. Examples, where this may be the case, include daily diary data (measures may include sleep quality, substance use, daily mood or affect) and data collected by loggers such as accelerometry measures of physical activity or logs of blood pressure. In these cases, researchers are interested in the true state of a behavior, but need to measure it multiple times to ensure they are capturing it with accuracy. Current state of the art practice with such data is to either average across the multiple observations or to create a latent factor from the multiple observations, which factor out the error and get a closer approximation of the true score. In repeated measures regression mixture models, the primary interest is in how such a factor or construct varies in its relation with other constructs while time is considered a nuisance rather than a factor of interest to explore. We propose that repeated measures regression mixtures will benefit by the information gained when using all available data rather than a summary score.

Regression mixture models are a specific type of finite mixtures, which examines heterogeneity in the regression of an outcome on predictor(s). Heterogeneity in effects is assessed by differences in the regression weights (e.g., intercept, slope), which connect predictor(s) to outcome(s) across latent classes. While a single set of average regression weights are estimated for all subjects in a traditional regression model, two or more sets of average regression weights are estimated for the corresponding number of latent classes in a regression mixture model. A general regression mixture model for a single outcome is:

$$y_{i|X,k} = \beta_{0k} + \sum_{p=1}^{P} \beta_{pk} \, x_{ip} + \varepsilon_{ik}, \qquad \varepsilon_{ik} \sim N(0,\sigma_k^2) \qquad \text{[eq. 1]},$$

where $y_{i|X,k}$ represents the observed value of y for subject i in a specific class k given predictor X, β_{0k} captures the class-specific intercept coefficient, and heterogeneity in the effects of each predictor is captured by β_{pk}, which represents the class-specific slope of the corresponding predictor, X_p. In line with traditional linear models, the errors within each class, ε_{ik}, are assumed to be normally distributed with a mean of zero and a class-specific variance of σ^2_k. Each observation (subject) is assigned to a latent class using a multinomial equation as a function of the overall latent class probabilities:

$$\Pr(c_i\!=\!k|\boldsymbol{z_i}) = \frac{\exp(\alpha_k + \sum_{q=1}^{Q} \gamma_{qk} z_{iq})}{\sum_{s=1}^{K} \exp(\alpha_s + \sum_{q=1}^{Q} \gamma_{qs} z_{iq})} \qquad \text{[eq. 2]},$$

where α_k denotes the log odds of being in class *k* versus the reference class when all latent class predictors, z_q (q=1,…, Q), equal zero. The potential difference in the predictor variable X_p between the latent classes can be taken into account by allowing the *C on X* path in the model (Lamont, Vermunt, & Van Horn, 2016). Figure 1 presents the example diagram of a regression mixture model with a single outcome (y), a single predictor (x), and where (c) represents the unobserved (latent) class membership.

When there are additional data for the outcome measure, we incorporate the framework of latent growth models (LGM; Duncan, Duncan, & Strycker, 2013) into the regression mixture models. While the current method can be extended to more complex models with additional time functions (e.g., linear-, quadratic-growth model), we focus on building a random intercept-only model for the repeated measures where inter-individual and intra-individual differences are adequately modeled. We specifically focus on detecting the heterogeneity in the effect of predictor(s) on the latent intercept. Adding additional latent factors for the time function (e.g., linear, quadratic slope factors) can be also considered after conducting more exploratoryanalyses, such as visual inspection using graphical function in statistical software.

Figure 2 represents the regression mixture model with 3 repeated measures of data. The single outcome regression mixture model can be extended for repeated measures as:

$$\mathbf{Y}_{i|X,k} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_{ik|X} + \boldsymbol{\varepsilon}_{ik}, \qquad \varepsilon_{tik} \sim N(0,\sigma^2_{tk}) \qquad \text{[eq. 3]},$$

where y_i refers to a vector of outcome variables (t×1, where t is the number of repeated measures from i individuals), τ refers to a vector of intercepts of ys (t×1; typically fixed to zero for model identification purpose), Λ relates the t observed repeated measures from the i individuals to the intercept (η_{ik}) for each latent class given the predictor X_p, and ε_{tik} represents the class-specific measurement errors of the repeated response variables (t×1; a vector of within-subject residual variance of ys for each class). In the simplest case, the Λ vector is fixed to be 1 at all measurement points to represent the average value of repeated outcomes across times and the variance-covariance matrix ∑ is a diagonal matrix with all diagonal elements fixed to be equal:

$$\mathrm{Var}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix} = \begin{bmatrix}\sigma_{1k}^2 & 0 & 0\\ 0 & \sigma_{1k}^2 & 0\\ 0 & 0 & \sigma_{1k}^2\end{bmatrix}$$

so that the residual variances are constant over time and uncorrelated for each pair of time measures (this is often called the Identity (ID) error structure). More complex error variance structures (e.g., Compound Symmetry, Auto-regressive, Banded-diagonal, etc.) could be considered. For the current study, however, we use the simplest ID structure for specifying the repeated measures regression mixture model as the method has limited value if it does not perform well under the simplest model. These restrictions presume that the focus of the model is on predicting the outcome equally across all repeated measures and that the residual variances across time points have no systematic correlations.

Given the measurement model for η in equation 3, the focus of the regression mixture is on the prediction of η as a function of the class specific effects of p covariates. Thus, η becomes the outcome in equation 1.

$$\eta_{ik|X} = \beta_{0k} + \sum_{p=1}^{P} \beta_{pk} x_{ip} + \zeta_{ik}, \qquad \zeta_{ik} \sim N(0, \psi_k) \qquad \text{[eq. 4].}$$

In equation 4, β_{0k} represents the mean of the repeated outcomes for each class when the predictor x_{ip} are zero, β_{pk} represents the class-specific regression coefficients of x_p as before, and ζ represents the deviation of the corresponding individual values from the mean estimates of the latent intercept factor. Along with the single latent intercept factor, there is one between-subject variance component (i.e., ψ_k), which captures the average deviation from the grand mean (i.e., mean of the repeated measures across all subjects) to the individual mean for the intercept factor.

While the current model is formulated under the LGM framework, it can be reproduced under the multilevel modeling (MLM) framework given the equivalence of the two model frameworks (Grimm, Ram, & Estabrook, 2017). When the model is reproduced in MLM, the 2-level model (i.e., repeated measures are nested within individuals) can be employed by having the total variance decomposed into ψ_k (i.e., between-subject variance) and σ^2_{tk} (i.e., within-subject variance). In the latent variable framework, this is achieved by having separate (correlated) variances across all of the measurement occasions.

Study aims

This study aims (1) to evaluate the performance of repeated measures regression mixtures and (2) to demonstrate the use of this model to examine the differential effects of stress on sleep diary in a sample obtained from a prior study. First, performance of the repeated measures regression mixture model is examined through Monte Carlo simulations in which we vary the measurement error variance of outcome across repeated measures, the number of repeated measures, and the sample size. In order to understand the usefulness of this approach in practice, we also compare the results of the repeated measures regression mixtures to two cases. The first case is a standard regression mixture where only 1 observation of the outcome is collected. This shows the potential benefit of collecting additional outcome data when planning a study, which will use regression mixtures. The second case compares the repeated measures regression mixture to the results that would be obtained if all repeated measures were available but the analysis conducted was a standard regression mixture with the outcome being the average across all time points. The purpose of this second case is to examine whether the benefits of using repeated measures data are due to simultaneously modeling multiple outcomes or to the increased reliability from having additional data points. If the two approaches (i.e., the repeated measures regression mixture approach and the single outcome with the average score regression mixture approach) show no difference in their improvement over the traditional single outcome approach, then we would conclude that the benefits are solely due to the increased reliability but not to simultaneously modeling multiple outcomes.

Outcomes in this study are both latent class enumeration (measured as the proportion of simulations that selected the correct number of classes) and the bias and efficiency of the parameter estimates. We hypothesize that both outcomes will be improved as the number of measurement and sample size increase. We also hypothesize that the benefit of having multiple outcomes is greater when the measurement error variance is larger, primarily due to jointly modeling these repeated measures outcomes. The reason for this is that the regression mixture model with a single outcome relies strongly on assumptions about the conditional distribution of that outcome for parameter estimation (Van Horn et al., 2015), with multiple outcomes, estimation is informed by the joint distribution of all outcomes, which contains more information than the univariate distribution.

We then demonstrate the use of repeated measures regression to identify qualitatively different groups of people who differ in the effect of cognitive and behavioral symptoms of depression on sleep efficiency. Sleep efficiency is measured using daily diaries, an example of the type of application, which may benefit from repeated measures regression mixtures. We believe that this proposed method is especially beneficial for the clinical research which often includes relatively small sample sizes and in which there is often an inherent interest in individual differences.

Methods

Data generation.

The first part of this study uses Monte Carlo simulations to examine the performance of the repeated measures regression mixture model. Data are generated using R (R Core Team, 2017) for a 2-class model as the true population model. Each class is defined by an intercept and slope (i.e., Class 1: $\beta_{00} = 0.0$ and $\beta_{10} = 0.20$ or 0.40 for the small or medium effect class, respectively; Class 2: $\beta_{01} = 0.50$ and $\beta_{11} = 0.70$ for the large effect class). Regression weights for Class 1 are set to be either small ($\beta_{10} = 0.20$) or medium ($\beta_{10} = 0.40$), so that the distances from Class 2 ($\beta_{11} = 0.70$) are set to be large- or small-distance condition, respectively. We anticipate that the regression mixture models perform poorly when the Class 1 effect size is medium given the reduced class separation from Class 2 (i.e., 0.40 vs. 0.70). We have used the current setting to generate the two latent classes because we believe that if regression mixtures are to be useful for finding individual differences they should be able to detect a difference of at least this size (Kim, Vermunt, Bakk, Jaki, & Van Horn, 2016; Van Horn et al., 2015). While the entropy is regarded as a model fit index to show the accuracy of class assignment in general finite mixture models, it is not necessarily indicative of goodness of fit in regression mixture models. Therefore, the entropy is not considered to be a design factor for our simulation study.

We generated 7 repeated measures outcome for each condition because collecting daily data for weekly basis is common in many applied research. Additionally, we vary the proportion of measurement error variance, σ^2_{1k} (1%, 10%, 20%, and 40% of the total variance[1]), ranged from minimal (1%) to substantial (40%). The reason we include the minimal measurement error condition is that we often see that there are variables rarely changed over time (e.g., daily Body Mass Index). When the measures are extremely stable and not changed, we expect that the benefits of having additional time points is much smaller than when there is some extent of

---

[1] For ψ_k, population values are set to be 99%, 90%, 80%, and 60%, accordingly).

measurement errors. These simulations include four different sample size conditions from 200 as a small sample size condition to 3000 as a large sample size condition. Although previous simulation study has shown that 3000 is an adequate number for detecting the differential effect with large effect size difference, many applied studies have employed the regression mixture approach with a much smaller sample size (Dyer et al., 2012; McDonald et al., 2016; Wong, Owen, & Shea, 2012). Therefore, we have included 200 as the smallest sample size condition in the current study. Sample sizes for the two classes are balanced, having 50% of the total samples assigned to each class (α_k=0.0). The predictor *X* was generated from a standard normal distribution with a mean of 0 and a standard deviation of 1. While it is important to model the potential mean difference in X between classes, we did not include this additional condition in the current simulations.

For data generation, there are a total of 32 separate conditions (2 effect size distance X 4 error variances X 4 sample sizes) with 500 replications used for each condition. The total amount of variance, $\beta_{1k}^2 + \sigma_{1k}^2 + \psi_k$, for each outcome is set to be 1 for all conditions. The following equations are used to generate the 7 time-point repeated measures data by varying the four conditions of measurement variance:

Class 1:

$$\begin{cases} \eta_{i|c=1} = 0.0 + 0.20x_i + \zeta_{i1}, & \zeta_{i1} \sim N(0, \begin{bmatrix} .950 \\ .864 \\ .768 \\ .576 \end{bmatrix}), & \varepsilon_{ti1} \sim N(0, \begin{bmatrix} .010 \\ .096 \\ .192 \\ .384 \end{bmatrix}), \\ \eta_{i|c=1} = 0.0 + 0.40x_i + \zeta_{i1}, & \zeta_{i1} \sim N(0, \begin{bmatrix} .832 \\ .756 \\ .672 \\ .504 \end{bmatrix}), & \varepsilon_{ti1} \sim N(0, \begin{bmatrix} .008 \\ .084 \\ .168 \\ .336 \end{bmatrix}), \end{cases}$$

Class 2:

$$\eta_{i|c=2} = 0.5 + 0.70x_i + \zeta_{i2}, \qquad \zeta_{i2} \sim N(0, \begin{bmatrix} .505 \\ .459 \\ .408 \\ .306 \end{bmatrix}), \qquad \varepsilon_{ti2} \sim N(0, \begin{bmatrix} .005 \\ .051 \\ .102 \\ .204 \end{bmatrix})$$

[eq. 5],

where ζ and ε are assumed to be normally distributed. The simplest variance-covariance (V-CV) structure is adopted for the within-subject residual V-CV (i.e., Identity structure)[2] because we specifically aim to use this approach for the repeated measures collected in a short period of time; we do not have a specific hypothesis for the error variance structure. Although it has been shown that the results of regression mixture models are affected by error variance distribution (Van Horn et al., 2012), examining non-normal errors is beyond the scope of the current study.

Data analysis.

Before analyzing the regression mixture models, a multiple regression model with a predictor X and a binary indicator of class membership, Z, (i.e., $y_i = \beta_0+\beta_1X+\beta_2Z+\beta_3X*Z+e_i$) is analyzed to confirm the data generation process. We used Mplus7.3 (L. K. Muthén & B.O. Muthén, 1998-2012) employing the maximum likelihood estimator with robust standard errors (ESTIMATOR=MLR) to analyze each dataset. Five types of regression mixture models are analyzed for each dataset: (A) a traditional regression mixture model with a single outcome measure (one of the 7 repeated measures), (B) 3-repeated-measures regression mixture model, (C) 5-repeated-measures regression mixture model, (D) 7-repeated-measures regression mixture model, and (E) model the average of all 7-repeated-measures in a regression mixture model. For

---

[2] This eventually makes the Compound Symmetry for the repeated outcome measures given the shared variance through the intercept variance, ψ.

Model (A), (B), and (C), the first 1, 3, and 5 variable(s) of the 7 generated outcomes are used for the data analysis, respectively. For Model E, the average score of the 7 repeated measures is used as a single outcome in a regression mixture analysis.

For all five specifications of regression mixture models, a series of analysis is conducted for each simulation varying the number of latent classes. First, a one-class model, which is a traditional regression model with no mixture, is analyzed. Results are compared with the next model having one more class (2-class model) using the Bayesian information criterion (BIC; (Schwarz, 1978) and the sample-size adjusted BIC (ABIC; Sclove, 1987). Finally, a three-class model was estimated for each simulated dataset to assess the probability of finding 3-classes over 2-classes. The number of simulations finding the correct number of classes (i.e., 2 classes) is reported in class enumeration results. Accuracy of parameter estimates is assessed using absolute bias in parameter estimates to the population values based on the following equation:

$$B(\hat{\beta}) = \overline{\hat{\beta}_{est}} - \beta_{pop} \qquad \text{[eq. 6].}$$

In addition, coverage rate for the true population value is also reported. More specifically, the coverage is coded as 1 if the true population value is included in the 95% confidence interval (CI); 0 if it is outside of the 95% CI. Only replications selecting the true 2-class model using the BIC are included for calculating the bias as well as the coverage rate of the estimated parameters[3]. To account for the problem of switched labels in simulations (McLachlan & Peel, 2000; Sperrin, Jaki, & Wit, 2010), we first sorted the two classes by the slope coefficients so that the smaller effect size class is always labeled Class 1 and the larger effect size class is Class 2.

[3] Full results including the parameter estimates for all simulation conditions are available from the first author.

Results

Class enumeration.

All 1-class models converged while convergence rate for the 2-class and 3-class models are ranged from 99.2% and 96.4%, respectively, to 100% depending on sample size. Table 1 presents the class enumeration results using the BIC[4]. Percentages of simulations detecting the true two classes are shown for each condition analyzed. Overall, the probability of selecting the correct 2-class model increases as sample size increases across all conditions as expected.

Adding additional time points. The primary question in the current study is whether the class enumeration result is improved by introducing the repeated measures into regression mixture model. This is assessed by comparing the results of the regular (single outcome) regression mixture (Model A) to the models with multiple repeated measures (Model B to Model D). As shown in Table 1, increasing the number of repeated measures in the regression mixture model substantially improves the probability of detecting the correct two classes using both the BIC and the ABIC. The average detection rate[5] of 28.9% for Model A using the BIC under the large distance condition is increased to 51.3%, 71.2%, and 82.8% for Model B, C, and D, respectively. Similarly, the detection rate under the small distance condition is gradually increased from 5.7% (Model A) to 52.9% (Model D) on average. Although the ABIC is shown to be less sensitive to the different model specifications, detection rate using the ABIC steadily increases as well by adding more repeated measures into the model[6].

---

[4] Class enumeration results using the ABIC can be requested from the first author.
[5] This is calculated using the sample size of 200, 500, and 1000, excluding 3000, across all conditions of measurement error variance.
[6] The results of the detection rate using the ABIC can be obtained by the request to the first author.

As hypothesized, the impact of adding additional time points on class enumeration is greater under the small sample size condition. For a sample size of 200 with the large distance condition, the probability of selecting the 2-class model using the BIC is substantially increased from Model A (average of 7.9%) to Model D (average of 52.3%) across different measurement error variance conditions. When the sample size is 500, Model D (7 repeated measures) selects the correct two classes in above 96% of simulations using the BIC while it is still below 20% for traditional regression mixtures (Model A) on average across all measurement error variance conditions. The class enumeration results for using the repeated measures are very promising when the sample size is 1000 under the large distance condition. The detection rate is 90% or above for using the three or more repeated measures in the model regardless of the degree of measurement error variance. While the same pattern is observed for the small distance condition, more repeated measures (at least five or seven) are required to achieve the power greater than .80. In sum, under the condition of small to moderately large sample sizes, introducing repeated measures into the regression mixtures greatly increases the utility of the method to detect the potential differential effects.

Using the average value of repeated measures. Next, we compare the class enumeration results of Model A and Model D to the results of an average composite model from the seven repeated measures (Model E). It is a common approach to use a composite variable of the repeated measures, such as mean and sum score, when the measures are expected to be stable with no systematic change over time. It is reasonable to expect that the average score represents the data better than the one-time measure at any point when the measurement error is associated with. Although there is an overall positive effect for using the average score over the single-time point measure on class enumeration results, the use of repeated measures (Model B, C, and D) outperforms the use of average score in regression mixtures using both the BIC and the ABIC.

While the detection rate is slightly increased from Model A to Model E by using the average score of 7 repeated measures, no major benefit has been shown for using the composite variable in regression mixture results. On the other hand, compared to Model E, using the original 7 repeated measure variables in regression mixture models (Model D) substantially improves the class enumeration results. Under the large distance between the two effect sizes condition, Table 1 shows that the average detection rate using the BIC is 28.9% for Model A and 39.1% for Model E while it is drastically increased to 82.8% for Model D. Similarly, the average detection rate is increased from 5.7% (Model A) and 9.7% (Model E) to 52.9% (Model D) by employing the repeated measures into the regression mixtures under the small distance condition.

Measurement error variance in repeated regression mixtures. As shown in Table 1, the degree of measurement error variance has no significant impact on class enumeration results for single outcome measure (Model A) other than the random sampling error. Given that the intercept variance and the residual variance are combined as one total residual variance component for Model A, it is not surprising that the class enumeration results are consistent under the same conditions with the total variance of one.

The interesting finding is that the impact of introducing the repeated measures into the regression mixtures is much greater when the measurement error variance gets larger. When the ratio of error variance is only 1%, the average rate for detecting the true 2 classes is increased from 29.2% to 45.7%, for Model A and B, respectively. Compared to 1% condition, when the measurement error variance is increased to be 10%, 20%, 40% of the total residual variance, the average detection rates increase to 48.4%, 52.1%, and 59.1%, respectively. That is, the impact of having more information from the repeated measures is greater when the measurement reliability is more problematic given the increased error variance. This can be understood as the more shared variance across repeated measures (higher ratio of intercept variance to the total variance[7]) adds less information on estimating the parameters. In the extreme scenario, when the ratio of intercept variance to the total variance is 1, the repeated measures are perfectly correlated and so they are redundant information. For Model C with a small or moderate sample size, the same pattern is shown that the detection rate steadily increases as the measurement error variance increases.

Distance in the effect size difference between classes. Under the condition of large distance in effect sizes between the two classes (.20 vs .70 condition), when the sample size is large (n=3000), two classes are successfully detected in most simulations regardless of the number of repeated measures as well as the degree of measurement error variance. Therefore, a sample size of 3000 condition is excluded for comparing different model specifications in class enumeration results for further discussion. Under the condition of small distance in effect sizes (.40 vs .70), even the sample size of 3000 condition cannot obtain the conventional power of .80 for successfully detecting the two latent classes using the traditional single outcome regression mixture model. This result is consistent with the previous simulation studies where 1-class model is selected over the true 2-class model in most simulations when the slope difference is small (Jaki et al., 2018). While it is not surprising that the overall coverage rate is relatively low in the small distance condition, the probability for selecting the 2-class model increases as the number of repeated measures increases. Since the class detection rate is noticeably low for the small

[7] This refers to the intra-class correlation (ICC) in the framework of multilevel modeling.

distance conditions, we further focus on the large distance condition for its accuracy of the parameter estimates.

Accuracy of parameter estimates.

Table 2 and 3 shows the summary of estimated parameters for a 2-class regression mixture model for all five types of model specifications[8]. Specifically, Table 2 shows the average bias in each estimated parameter (see equation 6) and Table 3 presents the coverage rate based on the 95% confidence interval (CI) constructed from its standard error. Simulations selecting the 2-class model as the best fitting model using the BIC[9] are included for computing the bias as well as the coverage rate. When sample size is 3000, biases for all parameter estimates are minimal across all simulation conditions[10]. This is consistent with the previous research that 3000 is an adequate number for detecting the true 2 classes under the current simulation condition (Jaki et al, in press). Therefore, we compare the absolute bias for the estimated parameters across five different model specifications when the sample size is small (n=200) to moderately large (n=1000). We focus on the parameters of the regression line (i.e., intercept and slope) for each class because the differential effects are mainly defined by the overall pattern of regression[11]. First, we describe the estimated parameters for Model A, compare the results to Model B through D, which use the repeated measures into the regression mixtures, and compare the results to Model E using the average score of 7 repeated measures.

---

[8] The model results for the small distance condition can be requested to the first author.

[9] Given that the BIC has shown more sensitive results on class enumeration, we used the BIC to summarize the accuracy of parameter estimates.

[10] We omitted the results for the sample size of 3000 and the measurement error variance of 10% in Table 2 to reduce the table size for a better presentation.

[11] The average estimate for factor variance and residual variance for all conditions can be also obtained by the first author on request.

In the following section, we focus on the results of Model A, which is the traditional single-outcome regression mixture model. As presented in Table 2, the average parameter estimates of Model A across different error variance conditions are severely biased when sample size is equal to or less than 1000. While the intercept coefficient and the regression weight for the large effect class (Class 2) are relatively well estimated across all conditions except for the sample size of 200, those for the small effect class (Class 1) are substantially biased depending on the model specification. When the population parameters for intercept and slope are β_{00}= .00 and β_{10}= .20 for Class 1 and β_{01}= .50 and β_{11}= .70 for Class 2, the average parameter bias across all sample size and measurement variance conditions, *B*, for the two estimates are $B(\beta_{00})$= -.20 and $B(\beta_{10})$= -.23 for Class 1 and $B(\beta_{01})$= -.01 and $B(\beta_{11})$= -.01 for Class 2, respectively. That is, parameter estimates for Class 1 are substantially downward biased while those for Class 2 are relatively stable. Class mean, which is the log of odds of being assigned to Class 1, is also severely biased to be $B(\alpha_0)$= -.65 (population class mean =.00). That is, about two times more subjects (66%) are assigned to be in Class 2, which dominates the class enumeration results.

The coverage rate also supports the findings from the bias in parameter estimates. Table 3 shows that the coverage rates for the population parameters under the small sample size condition is very low (ranged between .36 and .56) , meaning that the chances of recovering the true values are low for the sample size of 200, even when the two class model is selected as the best-fitting model. On the other hand, while the major biases in the parameter estimates are driven from the Class 1 (see Table 2), the coverage rate is relatively consistent between the two classes. The average coverage rate of 95% CI for intercepts and slopes for Class 1 are .72 and .74 while they are .77 and .71 for Class 2, respectively. The reason is that the standard errors for the estimated parameters for Class 1 are much larger than the ones for Class 2 given its instability and smaller class size.

Overall, results show that small sample size (n=200) is a serious problem in regular regression mixture models by producing biased parameter estimates, such as, switched direction for the regression weights (positive effect to be negative effect) and biased class proportion. Across all conditions of measurement error variance, the parameter estimates for Class 1 are downward biased (ranged between -.65 and -.23) and the power to detect the population parameter is low (ranged between .36 and .43).

Adding additional time points. When adding additional time points, accuracy of parameter estimates is substantially improved on average across all error variance conditions under the small to moderately large sample size conditions. The average biases for the intercept and slope for Class 1 are reduced to be $B(\beta_{00})$= -.09 and $B(\beta_{10})$= -.07 when adding two more time points (Model B). In a same pattern, when increasing the number of repeated measures to 5 and 7, the absolute biases for the intercept and slope for Class 1 steadily decrease to be $B(\beta_{00})$= -.05 and $B(\beta_{10})$= -.04 for Model C, and down to $B(\beta_{00})$= -.03 and $B(\beta_{10})$= -.02 for Model D, respectively. Conversely, the coverage rate of 95% CI to detect the population value increases as the number of repeated measures increases. Across all sample size and measurement variance conditions, the average coverage rates of all parameter estimates (β_{00}, β_{10}, β_{01}, β_{11}, and α_0) are increased gradually by adding more repeated measures into the regression mixture analyses. That is, for example, when a week of data (7 repeated measures) is used in regression mixture models instead of a single time point measure, biases in parameter estimates are dramatically reduced to be about 7 times smaller for intercept (-.20 to -.03) and 11 times smaller for regression weight (-.23 to -.02) on average, while the coverage rate is increased up to .95.

More specifically, under the condition of 20% measurement error variance with a sample size of 500, for example, the average estimate of the intercept and slope for Class 1 for Model A are β_{00}= -.17 (M_{SE} = .17) and β_{10}=.06 (M_{SE} = .14), which are seriously underestimated. That is, the statistical significance of small effect (population value of β_{10}= .20) cannot be even detected in most simulation conditions but regarded as 'no-effect' class. Notably, the standard deviation of the estimated regression weights for Class 1 across simulations is .30, which is substantial, indicating that the regression coefficient for the small effect class is unstable and estimated to be negative in many simulations. Compared to Model A, when using the 7 repeated measures in regression mixtures (Model D), the average regression weight for Class 1 under the same condition (n=500, error variance=20%) is β_{10}=.19 (M_{SE} = .09), which is very close to the population value. In terms of the statistical significance, it clearly shows the improved accuracy for estimating the regression weight for the smaller effect group with the average standard error of .09 when using the repeated measures approach. The standard deviation of the regression weights is .09 as well, which is greatly reduced from Model A (*SD*=.30) showing the improved model stability.

Likewise, the computed bias in class mean has been greatly reduced by increasing the number of repeated measures into the regression mixture models. When adding more repeated measures into the model, the average bias in class mean, $B(\alpha_0)$, is reduced to be -.19, -.08, and -.04 for Model B, C, and D, respectively, meaning that more subjects are correctly assigned to be in Class 1 given the increased number of repeated measures. This finding is also supported from the increased coverage rate as shown in Table 3.

Using the average value of repeated measures. Next, we examine whether using an average score of repeated measures has an advantage over a single time measure in the accuracy of parameter estimates. As shown in Table 2 and 3, Model E (the average score of 7 repeated measures model) shows no clear advantage over Model A overall. Although the size of biases for the intercept and slope for Class 1 ($B(\beta_{00})$= -.16 and $B(\beta_{10})$= -.16) are smaller than those from Model A ($B(\beta_{00})$= -.20 and $B(\beta_{10})$= -.23), the estimated parameters are still seriously biased and underestimated compared to Model B, C, and D. The absolute bias in class mean ($B(\alpha_0)$= -.43) and the coverage rate (.74) for Model E also show that the benefit of using the average score of the repeated measures is limited compared to the repeated measures regression mixture approach.

Application: Use of repeated measures regression mixture model to sleep research project data

To demonstrate the use of the repeated measures regression mixture model, we utilize data from the University of Memphis Sleep Research Project epidemiological survey of sleep and daytime functioning (Lichstein, Durrence, Riedel, Taylor, & Bush, 2004). While the overall association between the morbidity and sleep disturbances has been studied as a means of calibrating health risk (Altevogt & Colten, 2006; Nilsson, Rööst, Engström, Hedblad, & Berglund, 2004; Taylor et al., 2007), there has been a limited literature to understand the heterogeneity in the effects of daytime functioning on sleep quality, or vice versa. It is important to understand the potential differential effects because daytime functioning (e.g., depression, anxiety, and fatigue) affects the perceived sleep impairment, which has been identified as a significant factor in insomnia treatment-seeking behavior and increased use of healthcare (Moul et al., 2002; Ustinov et al., 2010).

In the current application study, regression mixture models are used to explore and better understand the potential differential effects of daytime functioning on sleep quality. Specifically, we illustrate the use of repeated measures regression mixtures to identify qualitatively different groups of people who differ in the effect of cognitive and behavioral symptoms of depression on sleep efficiency. A total of 771 participants ranged in age from 20 to 98 years participated in the study and were included for data analysis. More details on demographic and socio-economic characteristics of participants are available from Lichsten et al. (2004). Two weeks (14 days) of sleep diary were measured using a questionnaire. Among 9 sleep measures (e.g., time in minutes taking to fall asleep, the number of awakenings during the night, total sleep time, total sleep in bed, etc.), we employed the sleep efficiency percent for the current demonstration, which is the ratio of total sleep time to total time in bed x 100, indicating the overall sleep quality. Depression severity was measured using The Beck Depression Inventory (BDI) consisting of 21 items (Taylor et al., 2007).

To demonstrate the use of repeated measures regression mixture in the current application, we used the eight days of weekday data (Monday through Thursday for two weeks) to focus on the differential effect of daytime functioning on sleep quality during weekdays. Since previous research has shown that there are some different sleep patterns between the weekday and weekend (Adam, Snell, & Pendry, 2007; Vitale et al., 2014), we excluded the weekend data for the current illustration to remove the potential complications by combining the weekday and weekend data in the same analysis. First, to examine whether there is any systematic change of the sleep efficiency over time during weekdays, we plotted the individual trajectories across eight days of data from the randomly selected 15 subjects. Figure 3 shows no particular pattern in the repeated measures of sleep quality over time other than the random fluctuations. Next, we analyzed an intercept-only model under the MLM framework to assess the intra-class correlation coefficient (ICC), which is a gauge of the correlation among the observations within subjects (Raudenbush & Bryk, 2002). The calculated ICC is .49[12], indicating that about half of the variance is found at the between-subject level. We are specifically interested in understanding the between-subject variance part to the extent that we hypothesize the between-subject variance will be differentially explained by the daytime functioning (i.e., depression) based on the unobserved latent classes. Figure 4 shows the model specification for the repeated measures regression mixtures with sleep efficiency outcome having depression as the predictor. As shown in Figure 4, each sleep efficiency measure was treated as an indicator of a latent intercept factor of sleep efficiency with a fixed factor loading of 1, thus the latent variable represents the average sleep efficiency across eight repeated measures. The within-subject residual variances (σ^2_k) were constrained to be equal across time points but allowed to differ across classes and no covariances were estimated (i.e., Identity structure[13]). Intercept (β_{0k}), slope (β_{1k}), and between-subject residual variance (ψ_k) were estimated to be class-specific. A traditional regression analysis (1-class) was conducted and the results were compared to a series of regression mixture models (2-class, 3-class, 4-class, and 5-class). Mplus 7.3 (Muthén & Muthén, 1998-2012) was used for all data analyses and the BIC and ABIC were used for model comparisons[14].

[12] ICC = between-subject var/(between-subject var + within-subject var); .49 = 79.36/(79.36+81.10) from the results.

[13] To reduce the complexity of model specifications, the simplest ID error variance structure was adopted in the current study for demonstration purpose.

[14] Mplus syntax for analyzing the current model is provided in Appendix for applied researchers.

Table 4 presents the results of the repeated measures regression mixture model investigating the differential effect of depression on sleep impairment. Based on the BIC and ABIC for all four models, 4-class model fits the data best with the value of 42161 and 42100, respectively[15]. However, the Mplus result informs that the solution may not be trustworthy given that the best log-likelihood value was not replicated, which could be a sign of a local solution. Thus, we increased the random start value to be 500 (the Mplus default value is 20), which was suggested by the program. The result still warned that there was a non-convergence issue. Not surprisingly, 5-class model was not converged, which often is a sign that too many classes are requested to be extracted. Thus, we present both results from the 3-class and 4-class models for the readers of this paper (see Table 4) instead of presenting only one result over the other. Our interpretation on the model results is based on the 3-class model, which is converged with no issue.

As shown in Table 4, about half of the participants are classified to be in Class 1, named as a normal sleeper group, with 90% of sleep efficiency level on average with a small negative effect of depression. About 23% of the participants are classified to be in Class 2, named as a healthy sleeper group, with the highest sleep efficiency level on average (95.4%) with a smaller effect of depression on their sleep quality. In other words, participants in Class 2 are not greatly affected by the severity of depression on their sleep efficiency while they have overall a good sleep quality. About 28.4% of participants are classified to be in Class 3, named as a sleep impaired group, with the lowest level of sleep efficiency (79%) and the largest negative effect of depression on the quality of sleep. People in Class 3 are mostly the focus group for further

[15] We also analyzed the 5-class model, but the result was not converged, which can be a sign that we are trying to extract too many classes.

investigation in health and clinical research to find out who they are and why they are vulnerable to daytime functioning and psychological traits on their sleep. As shown in the current demonstration, regression mixture models allow applied researchers to investigate potential differential effects of interest on repeated measures outcome.

Discussion

Given the complexity of human nature, people respond differently to the same context. Regression mixture models are a powerful tool for identifying the unobserved class of subjects based on the differential effects of interest without a priori hypothesis. This study extends the use of traditional regression mixture models with a single observed outcome to a latent construct of repeated measures outcomes. The current approach for repeated measures regression mixture model can be considered as a special type of factor mixture modeling (Lubke & Muthén, 2005) where the population heterogeneity is lied on the effect of predictor X on the latent construct η, having the indicator Y as the repeated measures. While factor mixture modeling is a broader concept of latent class approach, which covers almost every type of population heterogeneity, repeated measures regression mixture models are specialized to detect the effect heterogeneity, which focuses more on the substantive questions.

While regression mixture models are a powerful tool for identifying the latent class showing the differential effects of interest, with only a single outcome the required sample size is very large and impractical in many applications (Jaki et al, in press; Van Horn et al., 2015). The current approach using repeated measures suggests an alternative solution by improving the power to search for the heterogeneous groups in the effect of interest. We note that this approach is especially recommended for those who are interested in the effect of heterogeneity with a small sample size when they have or are able to collect additional measurements from the same subjects. Collecting repeated measures data from the same subjects can be as difficult as increasing the sample size in many longitudinal studies, especially with measurements over longer period of time (e.g., attrition, relocation, mortality, etc.). The current method is particularly interested in using the repeated measures with relatively short time-interval, in which the main purpose of collecting the repeated measures is on achieving more stable measurements for the subjects rather than examining the change over time (e.g., daily diary of patients' pain severity, sleep diary, diary of physical activity, etc.). As shown in the results, introducing repeated measures into regression mixture models greatly improves the likelihood of identifying the correct number of latent classes, reduces bias in parameter estimates, and improves efficiency and stability of the results of analysis.

Whereas prior recommendations have suggested the regression mixtures require samples of well over 1000 participants even under the condition of large distance in effect sizes between the classes (Van Horn et al., 2015), the current repeated measures regression mixture models allow for samples as low as 200 participants when classes have separation like that simulated herein. Specifically, the results show that at the lowest sample size considered, 200, that more repeated measures are needed for problems like our simulation model; notable bias remained in the class with weaker regression effects even with 5 (but not 7) repeated measures. At samples closer to 500 results were reasonably stable and bias free when the large distance between two classes are conditioned. Thus, the use of repeated measures regression mixture modeling greatly extends the sample sizes for which regression mixtures are feasible, if there are repeated measures available. However, it is important to note that the results are dependent on including the repeated measures for specifying the regression mixture model rather than using the average score as a single outcome. Results have shown that the use of average score for the repeated measures in the regression mixture has no substantial but small gain in class enumeration accuracy or bias of parameter estimates. Another important point is that adding more repeated measures into the regression mixture models has a greater impact when the measures are relatively unstable. In other words, if the measures are highly stable across time, such as, repeated measures of Body Mass Index, the benefit of having the additional time point is limited given the small amount of extra information from the multiple time points.

Limitations. As in most simulation studies, there are some limitations regarding the simulation conditions in the current study. First, we have used two class as the population model in the current study. Because study conditions are already complex with 5 different model specifications for each simulated dataset, we have simplified the simulation conditions for generating the true population model. Although we have looked at a relatively small set of the possible regression mixtures, we believe that the major findings of the current study (i.e., improved power and accuracy for enumerating classes and estimating parameters in regression mixtures when using the repeated measures) will be held for extended conditions. Second, we elucidate this model using a relatively simple example where there is no time trend. This, of course, can be easily extended to the case where there is a time trend on the outcome. In this case, the repeated measures mixture is defined on the intercept term and/or the slope coefficient of the growth curve. Results of the current simulations should maintain in this slight extension, as long as the growth trend is modeled as a nuisance and constrained to be equal across classes. When the researcher's question is on the heterogeneity of growth trajectory, however, growth mixture model can be further considered for analyzing their data. We note that the substantive question is largely shifted from the differential effects to the differential trajectories in this case. More research would be needed to understand the case where instead a growth-mixture approach is needed or more appropriate.

We have purposefully constrained the total variance to be 1 for both latent classes across all simulation conditions to maintain the constant effect sizes regardless of the different conditions in measurement error variance. By doing this, increase in measurement error variance, which can possibly have a negative impact on the regression mixtures, is confound with the reduction in the intercept variance of the repeated measures, which can possibly have a positive impact on the results by providing more information for estimation from the multiple outcomes. Also, we used the simplest Identity V-CV for the residual error variance structure. It can be extended to more complex error variance structure for future research to examine whether the current approach still captures the latent classes appropriately and performs successfully under more complex scenarios.

For the application study, we simplified the model to be an unconditional regression mixture only with the main predictor, depression, without having any other covariates given the purpose of demonstration. We note that the results are intended to demonstrate the use of repeated measures regression mixtures rather than thoroughly examine heterogeneity in the effects of depression on sleep efficiency. Thus, findings of the current application study using the empirical data (i.e., three different types of sleepers given the effect of depression) should be interpreted with more thorough investigation.

Conclusion. Regression mixtures have been shown effective at finding differential effects in behavioral and health outcomes; however, large sample sizes are necessary to have stable and robust findings (Van Horn et al., 2015). Our repeated measures regression mixture approach holds extreme promise of reducing the size of the sample of individuals needed to use this methodology, as long as repeated measures are available. With the increase of repeated measures data associated with daily diary, logging and ecological momentary data, there are increasing opportunities to use this approach.

Open Practices Statement

None of the data or materials for the experiments reported here is available, and none of the experiments was preregistered.

References

Adam, E. K., Snell, E. K., & Pendry, P. (2007). Sleep timing and quantity in ecological and family context: A nationally representative time-diary study. Journal of Family Psychology, 21(1), 4-19. doi:10.1037/0893-3200.21.1.4

Altevogt, B. M., & Colten, H. R. (2006). *Sleep disorders and sleep deprivation: an unmet public health problem*: National Academies Press.

Jaki, T., Kim, M., Lamont, A. E., George, M., Chang, C., Feaster, D. J., Van Horn, M. L. (In press). The Effects of sample size on the estimation of regression mixture models. *Educational and Psychological Measurement.*

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application*: Routledge Academic.

Dyer, W. J., Pleck, J., & McBride, B. (2012). Using mixture regression to identify varying effects: A demonstration with parental incarceration. *Journal of Marriage and Family, 74*, 1129-1148.

George, M. R. W., Yang, N., Van Horn, M. L., Smith, J., Jaki, T., Feaster, D. J., & Maysn, K. (2013). Using regression mixture models with non-normal data: Examining an ordered polytomous approach. *Journal of Statistical Computation and Simulation, 83*(4), 757-770.

Grimm, K. J., Nilam, R., & Ryne, E. (2017). *Growth Modeling: Structural Equation and Multilevel Modeling Approaches.* New York: The Guilford Press.

Jung, T., & Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kim, M., Lamont, A. E., Jaki, T., Feaster, D., Howe, G., & Van Horn, M. L. (2016). Impact of an equality constraint on the class-specific residual variances in regression mixtures: A Monte Carlo simulation study. *Behavior research methods, 48*(2), 813-826.

Kim, M., Vermunt, J. K., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling, 23*(4), 601-614.

Kliegel, M., & Zimprich, D. (2005). Predictors of cognitive complaints in older adults: A mixture regression approach. *European Journal of Ageing, 2*(1), 13-23.

Kohli, N., Harring, J. R., & Zopluoglu, C. (2016). A finite mixture of nonlinear random coefficient models for continuous repeated measures data. Psychometrika, 81(3), 851–880.

Lamont, A. E., Vermunt, J. K., & Van Horn, M. L. (2016). Regression Mixture Models: Does Modeling the Covariance Between Independent Variables and Latent Classes Improve the Results?. Multivariate behavioral research, 51(1), 35-52.

DLanza, S. T., Cooper, B. R., & Bray, B. C. (2014). Population heterogeneity in the salience of multiple risk factors for adolescent delinquency. *Journal of Adolescent Health, 54*(3), 319-325.

Lanza, S. T., Kugler, K. C., & Mathur, C. (2011). Differential Effects for Sexual Risk Behavior: An Application of Finite Mixture Regression. *Open Family Studies Journal, 4*, 81-88.

Lee, E. J. (2013). Differential susceptibility to the effects of child temperament on maternal warmth and responsiveness. *Journal of Genetic Psychology: Research and Theory on Human Development, 174*(4), 429-449.

Lichstein, K. L., Durrence, H. H., Riedel, B. W., Taylor, D. J., & Bush, A. J. (2004). *Epidemiology of sleep: Age, gender, and ethnicity.* Mahwah, NJ: Erlbaum.

Lubke, G., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychologycial Methods, 10(1),* 21-39.

McDonald, S. E., Shin, S., Corona, R., Maternick, A., Graham-Bermann, S. A., Ascione, F. R., & Williams, J. H. (2016). Children exposed to intimate partner violence: identifying differential effects of family environment on children's trauma and psychopathology symptoms through regression mixture models. *Child abuse & neglect, 58*, 1-11.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Moul, D. E., Nofzinger, E. A., Pilkonis, P. A., Houck, P. R., Miewald, J. M., & Buysse, D. J. (2002). Symptom reports in severe chronic insomnia. *Sleep, 25*(5), 548-558.

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences*, 345-368.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research, 24*(6), 882-891.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (Seventh ed.). Los Angeles, CA: Muthén & Muthén.

Ng, S. K., McLachlan, G. J., Wang, K., Ben-Tovim Jones, L., & Ng, S. W. (2006). A mixture model with random-effects components for clustering correlated gene-expression profiles. *Bioinformatics, 22(14),* 1745-1752.

Nilsson, P. M., Rööst, M., Engström, G., Hedblad, B., & Berglund, G. (2004). Incidence of diabetes in middle-aged men is related to sleep disturbances. *Diabetes care, 27*(10), 2464-2469.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569.

Schmiege, S. J., & Bryan, A. D. (2016). Heterogeneity in the relationship of substance use to risky sexual behavior among justice-involved youth: a regression mixture modeling approach. *Aids and Behavior, 20*(4), 821-832.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Ann. Statist., 6*(2), 461-464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333-343.

Silinskas, G., Kiuru, N., Tolvanen, A., Niemi, P., Lerkkanen, M.-K., & Nurmi, J.-E. (2013). Maternal teaching of reading and children's reading skills in Grade 1: Patterns and predictors of positive and negative associations. *Learning and Individual Differences, 27*, 54-66.

Silinskas, G., Pakarinen, E., Niemi, P., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2016). The effectiveness of increased support in reading and its relationship to teachers' affect and children's motivation. *Learning and Individual Differences, 45*, 53-64.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Sperrin, M., Jaki, T., & Wit, E. (2010). Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics in Computing, 20*, 357-366.

Taylor, D. J., Mallory, L. J., Lichstein, K. L., Durrence, H. H., Riedel, B. W., & Bush, A. J. (2007). Comorbidity of chronic insomnia with medical problems. *Sleep, 30*(2), 213-218.

R Core Team. (2017). R: A language and environment for statistical computing. . doi: URL https://www.R-project.org/

Raudenbush, S.W.,&Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.).* Thousand Oaks, CA: Sage.

Ustinov, Y., Lichstein, K. L., Vander Wal, G. S., Taylor, D. J., Riedel, B. W., & Bush, A. J. (2010). Association between report of insomnia and daytime functioning. *Sleep Medicine, 11*(1), 65-68.

Van Horn, M. L., Jaki, T., Masyn, K., Howe, G., Feaster, D. J., Lamont, A. E., . . . Kim, M. (2015). Evaluating differential effects using regression interactions and regression mixture models. *Educational and Psychological Measurement, 75*(4), 677-714.

Van Horn, M. L., Smith, J., Fagan, A. A., Jaki, T., Feaster, D. J., Masyn, K., . . . Howe, G. (2012). Not quite normal: Consequences of violating the assumption of normality in regression mixture models. *Structural Equation Modeling, 19*(2), 227-249. doi: 10.1080/10705511.2012.65962

Vitale, J. A., Roveda, E., Montaruli, A., Galasso, L., Weydahl, A., Caumo, A., & Carandente, F. (2014). Chronotype influences activity circadian rhythm and sleep: Differences in sleep quality between weekdays and weekend. Chronobiology International, 32(3), 405-415. doi:10.3109/07420528.2014.986273

Wong, Y. J., Owen, J., & Shea, M. (2012). A latent class regression analysis of men's conformity to masculine norms and psychological distress. Journal of Counseling Psychology, 59, 176–183. doi:10.1037/a0026206

Xu, W. & Hedeker, D. (2001). A random-effects mixture model for classifying treatment response in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics, 11(4),* 253-273.

Yau, K. K., Lee, A. H., & Ng, A. S. (2003). Finite mixture regression model with random effects: Application to neonatal hospital length of stay. *Computational Statistics & Data Analysis, 41(3),* 359-366.