THE UNIVERSITY OF CHICAGO

EXTENSIONS TO STATISTICAL METHODS FOR ANALYZING INTENSIVE MULTILEVEL LONGITUDINAL DATA COLLECTED BY MOBILE AND WEARABLE DEVICES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PUBLIC HEALTH SCIENCES

BY
QIANHENG MA

CHICAGO, ILLINOIS
JUNE 2021

Dedicated to the mighty one, Lord Jesus Christ.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many thanks to brothers and sisters of the Church of Chicagoans, especially Stephen Hao and Benjamin Chen, Ted Teng, Courtney Wang, Xi Wang, who have fundamentally impact my life. It is amazing that we are able to grow together and fight together in Christ. Because of them, these past years mean much more than just completing a PhD degree. And I would also like to thank to my parents who have unconditionally supported me both financially and emotionally along the way. They just chose to trust every of my choices even when sometimes they couldn't understand the reasons behind.

I would forever be thankful to my PhD advisor, Professor Donald Hedeker for his kindness, patience and trust even when I was stuck and experienced hard times in doing research. It has been a wonderful and unforgettable experience of working with him. He is such a role model to me and I will never forget his insights and ever-lasting enthusiasm for research, which have generated long-lasting impact on my career. And I want thank Professor Robin J. Mermelstein from University of Illinois at Chicago and Professor Genevieve F. Dunton from University of Southern California for offering me such great opportunities to work with them. Experiences of working in their teams are eye-opening. And I would like to thank my dissertation committee members, including Professor Hedeker, Professor Mermelstein, Professor Yuan Ji and Professor Robert Gibbons. Lastly, I would like to thank Mrs. Michele Thompson, our program manager of Public Health Sciences, and Mrs. Siu-Chi Wong from UIC for the care and help I received from them.

At this end, I would like to mention those grants that financially supported the works in the thesis: the grant from National Cancer Institute (R01 CA240713) and the dissertation research grant from the Society of Multivariate Experimental Psychology, 2021.

ABSTRACT

Ecological Momentary Assessment (EMA) studies aim to explore how subjects' psychological states or behaviors interact with the real environment. Hand-held devices such as smartphones enable the participants in EMA studies to respond to the prompted assessments or self-initiate the assessments in real time. Besides conventional hand-held devices, wearable devices, e.g., actigraphy, have enabled more accurate and intensive tracking of subject's behaviors such as physical activities (PA). Relevant statistical methods for intensive longitudinal data analysis are continuing to be developed, however challenges remain such as modeling of the complex multilevel serial correlation in the data, informative nonresponses in self-initiated assessments, and identical data entries and also the irregularly distributed PA data. In this thesis, extended methods specifically for deal with the above mentioned issues were proposed.

First, during EMA studies, subjects can receive prompted assessments intensively across days and within each day, which results in three-level longitudinal data, e.g., subject-level (level-3), day-level nested in subject (level-2) and assessment-level nested in each day (level-1). Given the three-level EMA data, we proposed a linear mixed effects model with auto-correlated random effects at day-level and assessment-level. And with real time stamps of the assessments, we also provided a useful extension of this proposed model to deal with irregularly-spaced EMA assessments.

Second, we addressed the issue of non-responsivity of self-initiated assessments in EMA studies, where subjects are instructed to self-initiate reports when experiencing defined events, e.g., smoking. The frequency and determinants of non-responses in these event reports is usually unknown and these non-responses can even be associated with the primary longitudinal EMA outcome (e.g., mood) in which case a joint modeling of the non-responsitivity and

the mood outcome is possible. In certain EMA studies, random prompts, distinct from the self-initiated reports, may be converted to event reports, which provide some information about the subject's non-responsivity of event reporting. Using such data, we proposed a shared-parameter location-scale model to link the primary outcome model for mood and a model for subjects' non-responsivity by shared random effects which characterize a subject's mood change pattern and mood variability.

Third, for application of the MELS model, a problem that occurs is when subjects provide identical responses and therefore exhibit almost zero variance in their responses. It is assumed that certain latent clustering may exist to distinguish those who displayed different variance patterns. To deal with this, we assumed a mixture of normal distributions for scale random effects. For estimation, we incorporated Maximize-A-Posteriori (MAP) algorithm into the Expectation-Maximization (EM) algorithm framework so that we can estimate the posterior probability of clustering membership for each subject.

Lastly, to model the intensive irregularly distributed PA counts, we proposed a negative binomial mixed effects location-scale model (NBLS) to model these intensive longitudinal PA counts and to account for the heterogeneity in both the mean and dispersion level across subjects. Further, to handle the issue of inflated numbers of zeros in the PA data, we also proposed a hurdle/zero-inflated version which additionally includes the modeling of the probability of having non-zero PA levels.

CHAPTER 1

OVERVIEW

1.1 Mobile Health Technology

The growing use of smart electronic (wearable) devices and data technologies have given rise to the field of mobile health (mHealth) and significantly changed the landscape of public health research. First, these innovations have improved the reliability of measurements. Second, the convenience of data collection vastly increases data availability. In addition, due to the more frequent and timely exchange of information between researchers and participants, the obtained information is no longer static but instead dynamic, which displays more subtlety in the data. As a result, novel information technologies allow researchers to study daily life experiences of a single participant instead of limiting the scope on aggregate population features.

Studies and interventions regarding real-world experiences are becoming more popular in recent years. Conducting such studies is more economic and the attrition is less likely to occur because the time-span of the study is relatively short, e.g., for two or three weeks, compared to clinical trials which can last for a year or several years. In addition, these real-world studies have a more relaxed inclusion criterion so that the studied sample can be more representative for the population than clinical trials.

1.2 Ecological Momentary Assessments

In particular for psychological and behavioral studies, real-world data provide clues to uncover people's daily behavioral and psychological patterns that are closely related to the long-term health outcomes. To further design efficacious behavioral interventions or boosting strategies requires us to thoroughly understand how subject's psychological state and behavior are affected by transient environmental and contextual factors than long-term effects. As we mentioned above, the core of EMA study designs is the description or thorough assessments on the surrounding environment and participants' psychological states and how these two factors contribute to trigger momentary actions and behaviors.

EMA design also indicates a higher level of participants' involvement in determining the data availability and quality. For conventional EMA designs, the data availability completely relies on the participants' compliance and the way of collecting such data requires the input from the participants. For example, participants may receive the prompted surveys and then respond to the surveys, or initiate the input themselves via the devices. Another issue is data quality that the collected data may be subject to self-report bias. In order to resolve this, wearable devices can be used to collect the data in a automatically way, e.g., step counts recorded every 15 mins. And this way of data collection has also increased data availability and data reliability. But even with more advanced data collection technologies, how we deal with the observed subtlety as well as the irregular data distribution in such intensive longitudinal data is a still challenging topic .

1.3 Innovations of The Proposed Methods

As we have introduced above, practical issues may emerge in steps of data analysis, such as summarizing unusual data hierarchy and distribution, adjusting for bias in the data, dealing with missing information and efficient estimation when data size is large. Existing methods have not been sufficiently developed to deal with these issues, as will be shown in the following sections. In this dissertation, we proposed innovative extensions to statistical methods which can further serve as benchmarks to analyze other types of mobile health data. Below are the innovations of this thesis.

1.3.1 Utilizing the Real-Time Information

Data collected by mobile devices are often real-time stamped. Unlike conventional longitudinal studies where assessments happen every several months or every year, EMA data are usually intensively collected in a short time span and often irregularly spaced. Given this information, utilizing the time-stamp information with accuracy even to hours or minutes, will help us to more accurately adjust for the potential correlation among the assessments and reveal the environmental effects. In Chapter 2, we will propose a mixed effect model with hierarchical serial correlation structure and the extent of correlation can even depend on the length of real-time intervals.

1.3.2 Analyzing A More Challenging Type of EMA Data

Current methods for EMA data analyses are for data obtained from random prompts. However, few publications have touched the field of analyzing self-initiated event reports. These self-initiated event reports provide valuable clues to record subject's psychological state and other environmental momentary factors that trigger the event. For example, in such event reports, subjects can report their mood level and verify whether they are with others, when they are smoking. However, the main challenge of analyzing self-initiated event reports is that, the occurrence of nonresponses is unknown to the researchers. And inevitably, the data may be subject to bias coming from non-ignorable nonresponses, in other words, the non-responses are related to the outcome variable. Thus, in Chapter 3, we will propose a joint modelling approach which accounts for non-ignorable nonresponses and examines the association between responsivity and the event-contingent outcome variable for these self-initiated event reports.

1.3.3 Deal With Identical Responses Over Time

The identical responses over time can emerge in EMA surveys which assess participants' psychological states. The source of these identical entries can be mixed. Those can be the 'careless responses' potentially caused by the fatigue of responding to such intensive assessments but they can also be realistic entries. One way or another, those records will cause computational issues in estimating the within-subject variability. Current publications are less sufficient for handling within-subject duplicates. For the sake of computational convenience, those duplicates are considered to be redundant and then therefore removed. To avoid data exclusion, we can retain these subjects but assume they come from another cluster that display low within-subject variability. In Chapter 4, we will extend the mixed effect location-scale model (MELS)[25, 23] so that it can even be used to cluster different variance patterns for subjects.

1.3.4 New Estimation Approach for Generalized Linear Mixed Model

The proposed methods are extended from (MELS) , which belongs to the class of generalized linear mixed model (GLMM). GLMM is still the most accepted statistical method for inference of multilevel data. However, the complex model specification of MELS which accounts for the heterogeneity in variability is likely to be infeasible using conventional estimation approach. In Chapter 4 we will use another optimization algorithm to bypass the speed-limiting step in model estimation so that the model estimation will become much efficient.

1.3.5 Physical Activity Data Collected By Wearable Devices

With wearable devices such as smart watches or actigraphies, we can conveniently track subjects' activity over time. Those physical activity data are often non-negative valued counts data and irregularly distributed. For example, the data can display excessive amount of zeros since at most of the time subjects maintain sedentary. And the distribution of the counts can be severely right skewed because subjects often excessively exercise at just several time points. Although machine learning algorithms can be used to deal with these irregularly distributed time series and yield accurate prediction, a useful generalized regression model is still needed in terms of better interpretability. Thus, in Chapter 5, we propose a model which can deal with the irregularly distributed physical activity data and extract information regarding the subjects' activity patterns.

1.4 Summary

The rest of the dissertation is organized as follows. In Chapter 2, we will propose a linear mixed model with multilevel random effects that display complex serial correlation structure. In Chapter 3, to deal with the self-initiated event-contingent assessments, we will propose a joint modelling approach to relate the responsivity of the self-initiated assessments to the mean as well as the variability of these assessments. Then in Chapter 4, we will propose a mixed effect location-scale model with mixture-distributed random effects that provides information to classify subjects who are more/less likely to provide identical entries. In Chapter 5, to deal with the mentioned above issues in analyzing the physical activity counts, we propose a negative binomial mixed effect location-scale hurdle model to estimate latent subject-level effects in terms of mean, variability (dispersion) and the probability of having zeros. Lastly, in Chapter 6, we will summarize the proposed works.

CHAPTER 2

A THREE-LEVEL MIXED MODEL TO ACCOUNT FOR THE CORRELATION AT BOTH THE BETWEEN-DAY AND THE WITHIN-DAY LEVEL FOR ECOLOGICAL MOMENTARY ASSESSMENTS

2.1 Introduction

The Ecological Momentary Assessment (EMA) design [51, 26] is helpful to explore the interaction between subjects' psychological states and real environmental factors. In those EMA studies, psychologically or behaviorally related questionnaires are prompted to participants via hand-held devices multiple times within a day as well as across days. In other words, the collected EMA data are usually three-level data where assessments at level 1 are nested within days at level 2, and days are nested within subjects at level 3. Unlike other conventional longitudinal data where the interval between consecutive assessments is usually much longer (e.g., every six months), EMA assessments which are prompted in high frequency are usually longitudinally closely related to one another so that researchers may have to account for the potential correlation among assessments within the same subject. In addition, EMA assessments are usually unequal-spaced or called irregular-spaced as the EMA surveys are usually prompted randomly at any time of a day. These features of EMA data collection motivate us to develop a three-level model to account for the longitudinal correlation and to deal with the unequal-spacing.

For example, participants in the Adolescent Smoking Study [54] received random prompts to report their mood and the surrounding social context approximately five times per day for 7 days. In other words, there are at most 35 prompted assessments nested within a subject.

In this case, the repeated measures can be longitudinally correlated. Intuitively, correlation may exist between consecutive assessments within a day. Across days, the daily average of mood can also be inter-correlated due to subject's time-invariant personality traits and lifestyles that are unlikely to change over the short duration of a study. In particular, for negative mood (negative affect, *NEGAFF*), the primary outcome in this study, the between-day correlation among *NEGAFF* assessments can be endogenously related to certain subject's time-invariant personality traits.

Given such data, there can be two natural ways to model the within-subject correlation: (a) we ignore day as a level and allow for within-subject correlation of all the sequential assessments nested in a subject; (b) we allow level-1 assessment errors within a day to be correlated but assume uncorrelated level-2 random effects. The key assumption for (a) is that assessments across days are treated as the same as assessments within the same day. For a counter example, the two consecutive *NEGAFF* assessments across days may not be as correlated as the consecutive assessments within the same day because the overnight hours can interrupt the *NEGAFF* correlation. In particular for (a), we often assume the correlation intensity decays over the length of the assessment intervals. But in that case, correlation among across-day distant assessments can be underestimated to be almost zero. This is also a caveat of using approach (b), which assumes assessments across days are completely uncorrelated. Thus, we can argue that, in a realistic research setting, there exists a within-subject correlation structure that displays a correlation hierarchy where between-day (level-2) correlation and within-day (level-1) correlation are distinct and non-negligible. In the literature review below, we will show that most of the existing methods are basically the generalization of either the idea of (a) or (b) and they might be insufficient for multi-level EMA data analysis.

2.1.1 Two-level mixed effect model with random subject effects and autocorrelated errors

Mixed effect models [29] have been frequently used for handling heterogeneity across clusters in multi-level longitudinal data. For a two-level mixed model for repeated measures (level-1) nested within subject (level-2), the random subject effects summarize the latent subject-specific features apart from the observed covariates of interest. The subject random effects are usually assumed to follow a normal distribution with mean zero and are independent of the fixed effects and the random errors. Also, we often assume that the subject effects are independent across subjects. Instead of assuming this conditional independence among assessment errors, Chi and Reinsel [10] proposed a two-level mixed model including the so-called AR(1) structure (autocorrelation of order 1) as the error correlation structure. In their model, the variance-covariance matrix Σ_ϵ can further be expressed as $\Sigma_\epsilon = \sigma_\epsilon^2 \mathrm{P}$ where P is the correlation matrix, in this case, the AR(1) structure (Equation (2.1)).

$$\mathrm{P}_{AR(1)} = \begin{bmatrix} 1 & \rho & \dots & \rho^{n_i-1} \\ \rho & 1 & \dots & \rho^{n_i-2} \\ \vdots & & & \vdots \\ \rho^{n_i-1} & \rho^{n_i-2} & \dots & 1 \end{bmatrix} \tag{2.1}$$

For the AR(1) structure, the amount of correlation between two assessments within subject i conditional on fixed and random effects decays exponentially over their longitudinal distance. The correlation coefficient ρ is between 0 and 1. When no time-stamp information is provided, one way to quantify the longitudinal distance is to take the difference of the indices of the assessments, e.g., the correlation between assessment j_1 and assessment j_2 equals to $\rho^{|j_1-j_2|}$. This simple and intuitive form is beneficial for interpretation and computation, which makes AR(1) excessively used in times-series analysis.

When time-stamps are given, AR(1) can be extended as a serial AR(1) where the amount of correlation decays exponentially over the real longitudinal distance between two assessments, e.g., the correlation between assessment j_1 and assessment j_2 equals to $\rho^{|t_{j_1}-t_{j_2}|}$. Other forms of decreasing functions $h(|\ t_{j_1}-t_{j_2}\ |)$ can also be considered [59]. Equation (2.2) shows the serial power structure that we would use to represent the correlation coefficients in the extended model. This extended version helps to account for the unequal-spacing between consecutive within-subject assessments in EMA studies.

$$
\mathrm{P}_{SAR(1)} = \begin{bmatrix} 1 & \rho^{|t_2-t_1|} & \dots & \rho^{|t_{n_i}-t_1|} \\ \rho^{|t_1-t_2|} & 1 & \dots & \rho^{|t_{n_i}-t_2|} \\ \vdots & \vdots & & \vdots \\ \rho^{|t_1-t_{n_i}|} & \rho^{|t_2-t_{n_i}|} & \dots & 1 \end{bmatrix} \tag{2.2}
$$

While Chi and Reinsel [10] specified a correlation structure among the level-1 errors, Diggle proposed a decomposition for the level-1 error, e.g., $\epsilon_{ij} = c_{ij} + \epsilon_{ij}^{(0)}$ where c_{ij} is level-1 random effect with serial AR(1) autocorrelation [16]. This model with this decomposition is proved to be a more general form than Chi and Reinsel's model [59]. The term $\epsilon_{ij}^{(0)}$ represents the white noise of instrumental measurements in experiments such as blood test or chemical composition analyses but it is less meaningful for EMA self-reports where responses to assessments are subjective. And according to Diggle's comment [16], sometimes the data may not contain sufficient information to estimate both the subject random effects ν_i, the serial correlation c_{ij} and instrumental measurement errors $\epsilon_{ij}^{(0)}$. Therefore, even though Diggle's model is a more general form, for EMA data, we can still first follow Chi and Reinsel's model structure without this decomposition.

In general, these two-level models formalized idea (a) as mentioned. Even with autocorrelation among the within-subject assessments, they may still fail to account for the heterogeneity at day-level.

2.1.2 *Multi-level mixed models with level-1 autocorrelation*

Given this three-level or multi-level data structure, a mixed effect model with subject-level random effects ν_i and day-level random effects ψ_{ij} can be used to account for the subject-level and day-level heterogeneity. Extending Diggle's model for multi-level data, Vansteelandt and Verbeke even proposed a four-level mixed effect model but only level-1 random effects were assumed to be autocorrelated [57]. The EMA data they used were structured as four-level: subject-level (level-4), day-level nested in subject (level-3), signal-level (level-2) nested in a day and assessment-level for each signal (level-1). However, random effects at level 2, 3, 4 were assumed to be independent across units and their model didn't account for the possible correlation across signals (level-2) or across days (level-3). In addition, their model specified only intercept for fixed effects as they were more interested in the intraclass correlation than covariate effects. Their model actually formalizes idea (b) as mentioned. Again, it assumed uncorrelated random effects at the day-level, which made it inadequate to account for the autocorrelation across days in our scenario.

Other efforts to model level-1 autocorrelation include Anumendem et al. [3] who proposed a three-level mixed model with two separate serial AR(1) structures for level-1 units but again no level-2 autocorrelation was assumed. Their model combined the natural approaches (a) and (b) described in Introduction. It included not only the serial AR(1) for units within day but also an aggregate serial AR(1) for all units nested in each subject. Still, this model was not able to provide estimates for day-to-day autocorrelation.

2.1.3 Include lagged dependent variable as model covariates

Another intuitive way to account for the autocorrelation is to include the lagged dependent variable [2] as an extra explanatory variable to express the current dependent variable. As a result of doing this, biases together with inferential problems could be introduced when estimating fixed effects and the covariance components of random effects [2] as the assumption of statistical independence between random effects and fixed effects would be violated. For example, the time-invariant subject-level random effect also affects the lagged dependent variable of the same subject and in other words the lagged dependent variable added as a fixed effect could not be independent of the random subject effect.

Beyond the existing methods, we aim to propose a multi-level mixed effect model with separate AR(1) or serial AR(1) structures at both assessment-level (level 1) and day-level (level 2). We are going to validate the proposed model through a simulation study to compare our proposed models to other candidate models over 500 simulated datasets in terms of true parameter coverage and bias. And next, we will conduct the same set of model comparisons using the random prompts data in Adolescent Smoking Study using Akaike Information Criterion (AIC) and -2 log likelihood.

2.2 The proposed method

2.2.1 Data structure for the proposed model

The proposed model is for modelling three-level longitudinal data with day-level and assessment-level autocorrelation. The three-level data are structured as below:

Level 3, subject: $i = 1, 2 \ldots N$

Level 2, day: $j = 1, 2, \ldots, N_i$, nested within subject i;

Level 1, assessment: $k = 1, 2, \ldots, N_{ij}$, nested within day j within subject i.

In the real setting, the outcome of interest, *NEGAFF*, is usually treated as continuous normally distributed variable. The covariates can be time-invariant subject-level (level-3) characteristics, such as age, gender, level of education acquired prior to the study and subject's ability to deal with negative emotions. We will also include day-level covariates such as a binary indicator for whether the day of assessment is a workday/weekend. The assessment-level covariates usually reflect the transient contextual factors in the surrounding environment when the prompt was delivered, such as a binary indicator of whether the subject was alone or with other people when this subject completed the assessment.

2.2.2 The proposed model: RI_s+$AR(1)_d$+$AR(1)_{wd}$

Our proposed model is a three-level mixed effects model with independent subject-level random intercepts (RI_s) and AR(1) correlated day-level random effects ($\text{AR}(1)_d$) and AR(1) correlated within-day assessment errors ($\text{AR}(1)_{wd}$). The following sections show how we started from a simpler three-level mixed model structure and added on the level-1 and then the level-2 AR(1) correlation structures.

Three-level mixed effect model

For a three-level mixed effects model (Equation (2.3)), there is an independent random subject-level intercept ν_i to account for the heterogeneity across subjects and independent random day-level intercept ψ_{ij} to account for the heterogeneity across days. ϵ_{ijk} is the within-day error term. ν_i, ψ_{ij} and ϵ_{ijk} all follow zero-mean normal distributions but with different covariance structures. The variance of ν_i is σ_ν^2 and usually zero covariance is assumed between subjects. Both ψ_{ij}'s covariance matrix , Σ_ψ, and the error covariance matrix

Σ_ϵ can be either diagonal or with non-zero off-diagonal entries.

$$y_{ijk} = X_{ijk}^T\beta + \nu_i + \psi_{ij} + \epsilon_{ijk} \tag{2.3}$$

Level-1 (within-day) AR(1) structure

In each day j of each subject i, the random assessment errors (ϵ_{ijk}) follow a zero-mean normal distribution with the covariance matrix parameterized by a regular AR(1) correlation structure multiplied by a constant error variance σ_ϵ^2 (Equation (2.4)). The amount of correlation between consecutive assessments is assumed to be ρ_1. Again the regular AR(1) structure can be extended to the serial AR(1) version (Equation (2.2)) if time-stamped information is provided.

$$\begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij1} \\ \vdots \\ \epsilon_{ijN_{ij}} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1^{N_{ij}-1} \\ \rho_1 & 1 & \dots & \rho_1^{N_{ij}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{N_{ij}-1} & \rho_1^{N_{ij}-2} & \dots & 1 \end{pmatrix} \sigma_\epsilon^2 \right] \tag{2.4}$$

Level-2 (between-day) AR(1) structure

From now on, we add on another AR(1) structure to account for the between-day correlation for the day-level random intercepts within each subject (Equation (2.5)). For example, the amount of correlation between assessments of consecutive days is assumed to be ρ_2. The day-level correlation is a measure on how much the daily *NEGAFF* averages between two consecutive days are correlated.

$$\begin{pmatrix} \psi_{i1} \\ \psi_{i2} \\ \vdots \\ \psi_{iN_i} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2 & \dots & \rho_2^{N_i-1} \\ \rho_2 & 1 & \dots & \rho_2^{N_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_2^{N_i-1} & \rho_2^{N_i-2} & \dots & 1 \end{pmatrix} \sigma_\psi^2 \right] \tag{2.5}$$

The autocorrelation structures specified for level-2 random effects and level-1 errors are assumed to be sufficient for the multiple-level hierarchy of longitudinal correlation existing in the data.

2.3 Simulation setting

2.3.1 Model comparison

We compared our proposed model to the following candidate mixed effect models.

Model A: RI_s+$\text{AR}(1)_{ws}$

Model A is A two-level mixed effects model with random intercepts at subject-level (RI_s) and AR(1) within-subject correlation ($\text{AR}(1)_{ws}$) (Equation (2.6)).

$$y_{ik'} = X_{ik'}^T \beta + \nu_i + \epsilon_{ik'} \tag{2.6}$$

For this model, day-level is ignored and therefore the three-level data structure is re-organized as two-level: the between-subject level and the within-subject level. The within-subject assessments were re-indexed as $k' = 1, 2, \dots, \Sigma_j N_{ij}$. For example, if subjects in the study were followed up for 7 days and within each day subjects were instructed to complete 5 randomly prompted assessments, then it ended up with at most 35 within-subject assessments within each subject.

To account for the within-subject correlation, $\epsilon_{ik'}$ is assumed to follow a normal distribution with the variance-covariance matrix parameterized by a regular AR(1) correlation structure multiplied by a constant error variance (Equation (2.7)) for each subject i .

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i\Sigma_j N_{ij}} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \dots & \rho^{\Sigma_j N_{ij}-1} \\ \rho & 1 & \dots & \rho^{\Sigma_j N_{ij}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{\Sigma_j N_{ij}-1} & \rho^{\Sigma_j N_{ij}-2} & \dots & 1 \end{pmatrix} \sigma^2_{\epsilon'} \right] \tag{2.7}$$

The AR(1) structure can be replaced by serial AR(1) structure.

Model B: RI_s+RI_d+AR(1)_{wd}

Model B is a three-level mixed effects model (Equation (2.3)) with independent subject-level random intercepts (RI_s) and day-level random intercepts (RI_d) and AR(1) within-day correlation (AR(1)_{wd}) following the structure shown in Equation (2.4).

As these day-level random intercepts are assumed to be mutually independent within each subject, the level-2 correlation matrix is an identity matrix I_{N_i} of size N_i (Equation (2.8)).

$$\begin{pmatrix} \psi_{i1} \\ \psi_{i2} \\ \vdots \\ \psi_{iN_i} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad I_{N_i} \sigma^2_{\psi} \right] \tag{2.8}$$

The level-1 AR(1) (or serial AR(1)) structure for the level-1 random errors is the same as our proposed model.

Model C: with lagged dependent variables

Rather than to specify an AR(1) correlation structure for ψ_{ij} and ϵ_{ijk}, Model C used a naive approach to include the lagged dependent variables $y_{ij(k-1)}$ and $\overline{y_{i(j-1)}}$) to account for the autocorrelation (Equation (2.9)). $\overline{y_{i(j-1)}}$ here is the average of outcomes of the previous day $j-1$. In this model, the coefficients for the lagged dependent variables are assumed to be the correlation coefficients ρ_1 and ρ_2 at level-1 and level-2.

$$y_{ijk} = X_{ijk}^T\beta + \nu_i + \psi_{ijk} + \rho_1 y_{ij(k-1)} + \rho_2\overline{y_{i(j-1)}} + \epsilon_{ijk} \tag{2.9}$$

2.3.2 Data generation

For true values of the model parameters, for variances, we set $\sigma_\nu^2 = exp(1)$ and $\sigma_\psi^2 = exp(0.5)$ and $\sigma_\epsilon^2 = exp(0.5)$ and for correlation coefficients, we set $\rho_1 = 0.5$ and $\rho_2 = 0.25$. We transformed the estimates as well as the confidence intervals constructed from Wald's tests into log scale and logit scale so that the transformed confidence intervals were expected to be equal-tailed around the transformed estimates (Table 2.1).

We first generated 500 simulated datasets according to our proposed model. Within each dataset there are 100 subjects; nested within subjects there are 7 study days; within each day, there are 5 assessments. If the day-level is ignored, there are altogether 35 assessments within a subject.

The steps to simulate the data are as following:

1. We generated observed covariate at each level, e.g., x_{i3} is continuous covariate at level-3 (subject-level, L_3), x_{ij2} is the level-2 (day-level, L_2) covariate, and x_{ijk1} is the level-1(assessment-level, L_1) covariate.

2. We generated random independent subject effects ν_i according to the univariate normal distribution.

3. Given ρ_2 as the level-2 correlation coefficient, we used the multivariate normal distribution to generate the random day effects $\{\psi_{ij}\}$ according to Equation (2.5).

4. Within each day of each subject, given ρ_1 as the level-1 correlation coefficient, again we used the multivariate normal distribution to generate the random assessments according to Equation (2.4).

5. Then we simulated the outcome variable Y_{ijk} according to Equation (2.3).

After the data were generated, we then compared Models A, B, C with the proposed model on the bias of average of the model estimates from the true values, and on the coverage rate of each parameter. The coverage rate was computed as below:

$$Coverage = \frac{\text{number of times that the 95\% CIs covers the true value}}{\text{number of successfully convergent solutions}} \times 100\% \tag{2.10}$$

2.3.3 Comparing two AR(1) structures: regular AR(1) and serial AR(1)

Another simulation study was to compare the performances of the original proposed model with regular AR(1) correlated errors and the extended proposed model with serial AR(1) correlated errors.

We simulated the data from (i) the original model and (ii) the extended model. For scenario (i), we generated the data in the same way as Section 2.3.2 and after the data generation we appended to the dataset the random time-stamps so that these random time-stamps were not involved in the data generating process. In other words, the original proposed model in this case was the true model underlying the data. For scenario (ii), time-stamp information was used in generating the level-1 random errors so in this latter case the extended version of the proposed model was the true model underlying the simulated data.

To simulate the real time-stamps, we sampled the interval lengths from a uniform distribution $UNIF(0,1)$. For model comparison this time, we compared proposed models of the original version and the extended version based on the coverage rates as well as the biases to show the impacts of correlation structure mis-specification.

2.4 Estimation

2.4.1 Objective functions for optimization

Model A: RI_s+$\mathrm{AR}(1)_{ws}$

According to Section 2.3.1, given the ν_i, the conditional likelihood for the random error vector $\epsilon_i' = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{i(\Sigma_j N_{ij})})^T$ is,

$$f({\epsilon'}_i \mid \nu_i) = \frac{1}{\mid 2\pi\Sigma_{\epsilon'} \mid^{1/2}} exp(-\frac{1}{2}{\epsilon'}_i^T \Sigma_{{\epsilon'}_i}{\epsilon'}_i) \tag{2.11}$$

The form of the variance-covariance matrix Σ'_ϵ is shown in Equation (2.7). For estimating the extended version of this model, we replaced the regular AR(1) structure by serial AR(1) structure as shown in Equation (2.2).

The distribution of ν_i (a scalar) is,

$$g_3(\nu_i) = \frac{1}{\sqrt{(2\pi\sigma_\nu^2)}} exp(-\frac{\nu_i^2}{2\sigma_\nu^2}) \tag{2.12}$$

So the marginal likelihood for ϵ'_i is,

$$\mathcal{L}(\epsilon') = \int \prod_{i=1}^{N_i} \{f(\epsilon'_i \mid \nu_i) g_3(\nu_i) d\nu_i\} \tag{2.13}$$

Model B: RI$_s$+RI$_d$+AR(1)$_{wd}$

Given the independent random day effects $\{\psi_{ij}\}$ and random subject effects $\{\nu_i\}$, the conditional likelihood for the random error vector $\epsilon_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \ldots, \epsilon_{ijN_{ij}})^T$.

$$f(\epsilon_{ij} \mid \nu_i, \psi_{ij}) = \frac{1}{\mid 2\pi\Sigma_\epsilon \mid^{1/2}} exp(-\frac{1}{2}\epsilon_{ij}^T \Sigma_\epsilon \epsilon_{ij}) \tag{2.14}$$

The distribution of $\psi_i = (\psi_{i1}, \psi_{i2}, \ldots, \psi_{iN_i})^T$is,

$$g_2(\psi_i) = \prod_{j=1} \frac{1}{\sqrt{2\pi\sigma_\psi^2}} exp(-\frac{\psi_{ij}^2}{2\sigma_\psi^2}) \tag{2.15}$$

The distribution of ν_i (a scalar) is,

$$g_3(\nu_i) = \frac{1}{\sqrt{(2\pi\sigma_\nu^2)}} exp(-\frac{\nu_i^2}{2\sigma_\nu^2}). \tag{2.16}$$

So the marginal likelihood for optimization is given as,

$$\mathcal{L}(\epsilon) = \int_\nu \int_\psi \prod_{i=1}^{N} \{ \prod_{j=1}^{N_i} \{ f(\epsilon_{ij} \mid \nu_i, \psi_{ij}) g_2(\psi_{ij}) d\psi_{ij} \} g_3(\nu_i) d\nu_i \}. \tag{2.17}$$

Again, in order to estimate the extended version of Model B, we replaced the regular AR(1) structure by the serial AR(1) structure at within-day level.

Model C: with lagged dependent variables

We followed the same approach to estimate this model as we did for Model B. One issue is that for the first assessment in each day or for the first day within a subject there is no day-level and assessment-level lagged dependent variables. We adopted a technique described by Jones [27] to prevent these observations to be deleted during programming so that we could still use all the available data.

Proposed model: RI_s+$\text{AR}(1)_d$+$\text{AR}(1)_{wd}$

Given ψ_{ij} and ν_i, the conditional likelihood for the random error vector will still follow Equation (2.14) but the variance-covariance matrix Σ_ϵ will be in the AR(1) form (Equation (2.4)). For estimating the extended version of our proposed model, as we did before, we just replaced the regular AR(1) structure in the variance-covariance matrix of the random errors by the serial AR(1) structure.

For day-level random effects, the distribution of $\psi_i = (\psi_{i1}, \psi_{i2}, \ldots, \psi_{iN_i})^T$ is,

$$g_2(\psi_i) = \frac{1}{| 2\pi\Sigma_\psi |^{1/2}} exp(-\frac{1}{2}\psi_i^T \Sigma_\psi \psi_i) \tag{2.18}$$

And the variance-covariance matrix of the day-level random effects is specified by AR(1) structure (Equation (2.5)).

The distribution of ν_i (a scalar) is the same as Equation (2.12). So the objective function for optimization is given as,

$$\mathcal{L}(\epsilon) = \int_\nu \int_\psi \prod_{i=1}^{N} \{f(\epsilon_i \mid \nu_i, \psi_i) g_2(\psi_i) g_3(\nu_i) d\psi_i d\nu_i\} \tag{2.19}$$

2.4.2 Software and optimization

Conventionally, we used maximum marginal likelihood estimation (MMLE) to estimate the generalized linear mixed model (GLMM). Prior to optimization, the marginal likelihood function is evaluated by integrating the joint log-likelihood function over the distribution of the random effects. Numerical approximation methods such as adaptive Gaussian quadrature [44] or Laplace approximation are necessary. This is a more general approach for estimating arbitrary GLMM without utilizing distributional assumptions. But under special circumstances, for example, for normally distributed responses or errors, we can directly derive closed form solution for the marginal likelihood, in which we can bypass the complex numerical integration to expedite estimation.

We used PROC GLIMMIX of SAS 9.4 (SAS Institute Inc., Cary, NC) for the estimation.

Primarily, Newton-Raphson Method was used for the optimization. The estimation method used in PROC GLIMMIX is the maximum pseudo likelihood (MPL) method. The MPL method uses Taylor expansion to linearize the non-linear relationship between the response and the fixed and random effects and then produce asymptotic normal 'pseudo responses'. Given these asymptotic normal responses, as we have discussed above, we can directly yield closed-form solutions for the fixed and random effects. For more general situations, likelihood constructed from the pseudo responses is not the true likelihood, making model comparison illegitimate. However, when the responses and errors are normally distributed and identity link function is used, the MPL method is equivalent to the MMLE method. In our case, as we are modelling normal responses, estimates given by MPL are the same as those from MMLE. In PROC GLMMIX, we specified NOREML option in order to obtain the value of the likelihood function with unadjusted degree of freedom. Other benefit of using PROC GLIMMIX is the convenient specification of hierarchical correlation structure regardless of the situation where some days and some within-day assessments might be missing.

2.5 Results of simulation study

2.5.1 Comparing different model structures

Table 2.1 shows the comparison on the means and standard deviations (SD) as well as the coverage rates (CR) across four candidate models. Data used in this simulation study was simulated according to our proposed model in 2.4.1 and our proposed model successfully recovered all model estimates with reasonable biases and coverage rate around 0.95.

Even though Model A, B and the proposed model had different random effect distributions, they still yielded similar means and standard deviations on the β estimates because their random effect distributions were independent of β. However, Model C adopted a different

Table 2.1: Compare model A, B, C and the proposed model on 500 datasets simulated from the original proposed model with regular AR(1) errors

	Model A		**Model B**		**Model C**		**Proposed Model**		
	RI_s+$AR(1)_{ws}$		**RI_s+RI_d+$AR(1)_{wd}$**		**w/ Lagged DVs**		**RI_s+$AR(1)_d$+$AR(1)_{wd}$**		**True**
	CR	**Mean(SD)**	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**Value**
$\boldsymbol{\beta_{Int}}$	0.950	4.985(0.402)	0.949	4.983(0.396)	0.190	3.932(0.360)	0.954	4.972(0.373)	5.0
$\boldsymbol{\beta_{L1}}$	0.923	0.400(0.011)	0.945	0.400(0.010)	0.860	0.409(0.011)	0.948	0.400(0.009)	0.4
$\boldsymbol{\beta_{L2}}$	0.757	-1.004(0.065)	0.937	-1.004(0.061)	0.648	-0.914(0.059)	0.934	-1.010(0.062)	-1.0
$\boldsymbol{\beta_{L3}}$	0.945	0.700(0.178)	0.941	0.701(0.177)	0.874	0.585(0.151)	0.942	0.702(0.168)	0.7
$\boldsymbol{log(\sigma_\nu^2)}$	0.961	0.999(0.159)	0.954	1.017(0.156)	0.550	0.633(0.190)	0.966	0.959(0.168)	1
$\boldsymbol{log(\sigma_\psi^2)}$			0.881	0.400(0.101)	0.920	0.466(0.077)	0.952	0.497(0.108)	0.5
$\boldsymbol{log(\sigma_\epsilon^2)}$	0.000	1.163(0.047)	0.962	0.499(0.055)	0.000	0.152(0.030)	0.964	0.499(0.054)	0.5
$\boldsymbol{logit(\rho_1)}$	0.000	0.528(0.066)	0.964	-0.003(0.113)	0.000	-1.957(0.142)	0.944	-0.004(0.113)	0
$\boldsymbol{logit(\rho_2)}$					0.000	-2.571(0.555)	0.960	-1.160(0.509)	-1.1

RI_s: independent subject-level random intercept(RI)
$AR(1)_{ws}$: AR(1) correlated within-subject(ws) random errors
RI_d: independent day-level random intercept(RI)
$AR(1)_d$: AR(1) correlated random day effects (level-2)
$AR(1)_{wd}$: AR(1) correlated within-day(wd) random errors (level-1)
CR: coverage rate
SD: standard deviation

approach to account for the autocorrelation at level-1 and level-2. Instead of assuming correlation among the random effects, Model C augmented the fixed effects by including the level-1 lagged outcome and the lagged average of the outcomes at level-2, i.e., the daily average of the previous day. As we have discussed, the lagged variables were not independent of the random effects, which resulted in lower coverage rate on level-3 fixed effect estimates and non-ignorable biases for covariance parameters at all three levels.

With the same set of fixed effects, three-level models (Model B and the proposed model) fit the simulated three-level data better than the two-level model (Model A). Using a two-level model on the three-level data may have low coverage on level-2 fixed effect and therefore inferential issues can occur. While comparing two three-level models, using a three-level mixed effect model but assuming independent level-2 random intercepts can bring in bias for the level-2 random effect variance. Thus, using the wrong three-level model may further lead to a wrong conclusion on the intraclass-correlation-coefficient (ICC). In other words, we may not able to correctly identify the source of variation.

This part of results of simulation showed the adverse effects of model mis-specification on the data simulated from our proposed model. All other candidate models, Model A, B and C were shown to have fatal shortcomings in estimation or inference.

2.5.2 Comparing proposed models with AR(1) vs. with serial AR(1) structure

Table 2.2 shows the results of assuming regular AR(1) errors on data generated from the extended version with serial AR(1) errors. Given the data simulated from the extended proposed model, both the original version and the extended version yielded similar results over β. The extended version, the true model to generate the data, had reasonable coverage rate of around 95% and the means of 500 solutions were close to the true values of the model parameters. However, using the original version with regular AR(1) structure can lead to significant bias and low coverage on the autocorrelation coefficient. We also compared the AIC values as well as the -2 Log Likelihood values between the two versions (original vs. extended). On these data simulated from the extended version, if we selected model with lower AIC or -2 Log Likelihood, the correctness of decision making was 97.4% based on either AIC or -2 Log Likelihood.

Conversely, for data simulated from the original version, the correctness of decision making based on smaller AIC and -2 Log Likelihood was 100% among 500 simulated datasets. For these data, the original version, the true model for data generation, yielded little bias and the resulted confidence intervals had coverage of close to 95% on the parameters. However, using extended version on these data can result in slightly lower coverage on level-1 and level-2 fixed effects and significant bias and low coverage for level-1 and level-2 variances as well as the correlation coefficients (Table 2.3). Thus, by using the extended version on data generated from the original version, inference for both the level-1 and level-2 fixed effects

Table 2.2: Model Comparison over 500 datasets simulated from the extended proposed model with serial AR(1) errors

	Original Version		**Extended Version**		
	$\mathbf{RI}_s$**+AR(1)**$_d$**+AR(1)**$_{wd}$		$\mathbf{RI}_s$**+AR(1)**$_d$**+SAR(1)**$_{wd}$		**True**
Parameters	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**Value**
$\boldsymbol{\beta_{Int}}$	0.954	4.937(0.762)	0.954	4.938(0.761)	5.0
$\boldsymbol{\beta_{L1}}$	0.950	0.399(0.010)	0.946	0.399(0.010)	0.4
$\boldsymbol{\beta_{L2}}$	0.948	-1.005 (0.118)	0.948	-1.005(0.118)	-1.0
$\boldsymbol{\beta_{L3}}$	0.956	0.734(0.365)	0.958	0.734(0.365)	0.7
$\boldsymbol{log(\sigma_{\nu}^2)}$	0.938	0.960(0.172)	0.938	0.960(0.172)	1.0
$\boldsymbol{log(\sigma_{\psi}^2)}$	0.964	0.497(0.093)	0.968	0.497(0.092)	0.5
$\boldsymbol{log(\sigma_{\epsilon}^2)}$	0.952	0.501(0.043)	0.950	0.500 (0.042)	0.5
$\boldsymbol{logit(\rho_1)}$	0.004	-0.579 (0.121)	0.968	-0.006(0.104)	0.0
$\boldsymbol{logit(\rho_2)}$	0.964	-1.184(0.453)	0.946	-1.185(0.452)	-1.1

RI_s: independent subject-level random intercept(RI).
$\mathrm{AR}(1)_d$: AR(1) correlated random day effects (level-2).
$\mathrm{SAR}(1)_{wd}$: Serial-AR(1) correlated within-day(wd) random errors.

and variances can be problematic.

2.6 Application on Adolescent Smoking Study data

2.6.1 The Adolescent Smoking Study data

There were 461 subjects in the Adolescent Smoking Study [54]. During the study, subjects were instructed to complete approximately 5 assessments prompted by electronic devices per day for the first 7 days. We compared the goodness-of-fit (AIC, -2 log likelihood) of three candidate models (Model A, B and the proposed model) to see whether there is evidence of the level-1 and level-2 correlation hierarchy we elaborated in 2.2.2. The outcome of interest, *NEGAFF*, is an average over 5 items and each item rated from 1-10. Altogether 12,059 observations were used in this analysis.

The time-stamps of responses were recorded in minutes from 0 a.m. in each day (0 min to 1440 mins) and later divided by 60 to generate time-stamps in unit of 1 hour. We then ad-

Table 2.3: Model comparison over 500 datasets simulated from the original proposed model with serial AR(1) errors

	Original Version		**Extended Version**		
	RI_s+AR(1)$_d$+AR(1)$_{wd}$		**RI_s+AR(1)$_d$+SAR(1)$_{wd}$**		**True**
Parameters	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**Value**
$\boldsymbol{\beta_{Int}}$	0.954	4.986(0.395)	0.921	4.994(0.392)	5.0
$\boldsymbol{\beta_{L1}}$	0.948	0.400(0.010)	0.903	0.400(0.012)	0.4
$\boldsymbol{\beta_{L2}}$	0.934	-1.004 (0.061)	0.927	-1.004(0.062)	-1.0
$\boldsymbol{\beta_{L3}}$	0.942	0.701(0.176)	0.951	0.699(0.174)	0.7
$\boldsymbol{log(\sigma_{\nu}^2)}$	0.966	0.959(0.168)	0.968	0.963 (0.165)	1.0
$\boldsymbol{log(\sigma_{\psi}^2)}$	0.952	0.497(0.108)	0.324	0.687(0.080)	0.5
$\boldsymbol{log(\sigma_{\epsilon}^2)}$	0.964	0.499(0.054)	0.005	0.272(0.052)	0.5
$\boldsymbol{logit(\rho_1)}$	0.960	-0.004(0.113)	0.000	-1.940 (2.218)	0.0
$\boldsymbol{logit(\rho_2)}$	0.944	-1.160(0.509)	0.854	-1.489 (0.454)	-1.1

RI_s: independent subject-level random intercept(RI).
AR(1)$_d$: AR(1) correlated random day effects (level-2).
SAR(1)$_{wd}$: Serial-AR(1) correlated within-day(wd) random errors.

justed the day ID according to the time stamps: we assumed the first 3 hours of each day (0-3 am) as the last 3 hours continuing from the day before. As we have stated in the motivating example in Section 2.1, participants responded to the random prompts approximately five times per day for 7 days and therefore completed at most 35 prompted assessments within subject. Data for this application were unbalanced and missing data existed.

Covariates included in the model were of three different levels. Subject-level (level-3, L_3) covariates include *gender* (male vs. female), *grade10* (whether the participant completed 10th grade (=1) or 9th grade (=0)), *NovSeek* (the ability of novelty seeking) and *NegMoodReg*(the ability to regulate negative mood). For day-level(level-2, L_2), we included *Weekend* (whether the day of assessment is weekend day (=1) or not (=0)) in to the model. And for assessments-level(level-1, L_1), we had *ALONE* (whether a subject is alone (=1) or not (=0) when completing this assessment). We further decomposed the *Alone* variable into *AloneBS* and *AloneWS*, e.g., *Alone=AloneBS+AloneWS*. *AloneBS* is the within-subject average of *Alone* and the corresponding β coefficient reflects how likely a participant to be alone and it was assumed to be a time-invariant subject-level characteristic. *AloneWS* can be viewed as mean-centeringed *Alone* and its effect can be interpreted as the momentary effect of being alone at the moment of assessment. By this decomposition, the effect of the subject characteristic *AloneBS* and the effect of the momentary contextual variable *AloneWS* were allowed be different.

2.6.2 Results

From Table 2.4 and Table 2.5, all six candidate models yielded similar estimate and significance level on fixed effect β, no matter for the original versions or their extended versions. Males had more negative mood then females. Participants who completed the 10th grade had more negative mood than participants who hadn't. Novelty seeking (*NovSeek*) and negative mood regulation (*NegMoodReg*) represents participants' ability to deal with adverse emotions and both of them were significant. The negative effect of *NegMoodReg* is intuitive that participants with better ability to regulate negative mood had less negative mood. However, if participants had higher score of novelty seeking might experience more negative mood. *AloneBS* and *AloneWS* were both significant and associated with more negative mood. *Weekend* was the only day-level covariate in the model and *NEGAFF* was significantly less negative at weekends versus at weekdays.

Comparing across three non-extended versions of the candidate models, our proposed model yielded the best fit on the data with the lower AIC and -2 loglikelihood values than Model A, B. Although all three models accounted for the correlation among within-subject assessments, the three-level model hierarchy was proved to perform significantly better than the two-level model if given the three-level data structure. The subject variance of Model A is the largest ($\sigma_\nu^2 = 1.685$). As we introduced more sources of variation, the subject variance displayed a decreasing trend from Model A to Model B ($\sigma_\nu^2 = 1.673$) to our proposed

Table 2.4: Model comparison on data from Adolescent Smoking Study (original versions)

Parameters	Model A $\mathbf{RI}_s+\mathbf{AR(1)}_{ws}$	Model B $\mathbf{RI}_s+\mathbf{RI}_d+\mathbf{AR(1)}_{wd}$	Proposed Model $\mathbf{RI}_s+\mathbf{AR(1)}_d+\mathbf{AR(1)}_{wd}$
$\beta_{Intercept}$	5.212***	5.214***	5.105***
$\beta_{Male(vs.female)}(L3)$	-0.384**	-0.388**	-0.381**
$\beta_{Grade10}(L3)$	0.088	0.093	0.087
$\beta_{NovSeek}(L3)$	0.216*	0.216*	0.235*
$\beta_{NegMoodReg}(L3)$	-0.807***	-0.806***	-0.792***
$\beta_{AloneBS}(L3)$	0.958**	0.968**	0.950**
$\beta_{AloneWS}(L1)$	0.374***	0.362***	0.363***
$\beta_{Weekend}(L2)$	-0.207***	-0.246***	-0.221***
σ^2_ν	1.685	1.673	0.747
σ^2_ψ		0.454	1.439
σ^2_ϵ	2.918	2.465	2.423
ρ_1	0.318	0.206	0.193
ρ_2			0.835
$-2loglik$	47083	47044	46883
AIC	47105	47068	46909

*** p-value $<$ 0.0001; ** p-value $<$ 0.01; * p-value $<$ 0.05;

model ($\sigma^2_\nu = 0.747$). Especially after we accounted for the day-to-day correlation, σ^2_ν decreased drastically from Model B to the proposed model. This decrease of subject variance indicated that the day-level variance and then the day-to-day autocorrelation effectively explained part of the subject-to-subject variation. In addition, the autocorrelation intensity of Model A ($\rho_1 = 0.318$) was between the level-1 autocorrelation ($\rho_1 = 0.193$) and level-2 autocorrelation ($\rho_2 = 0.835$) in the proposed model, which indicated that the within-subject autocorrelation in Model A summarized the correlation at assessment level as well as at day level.

Comparing the intensity of autocorrelation between level 1 and level 2 in the proposed model, we observed that level-2 between-day autocorrelation intensity is much stronger than the level-1 within-day autocorrelation. As the between-day autocorrelation quantifies the correlation between daily averages of *NEGAFF* assessments, this strong intensity comes from the consistency of subjects' personality trait or routine lifestyle. And the assessment-level variance was the largest among variances of three levels. In other words, even after controlling for the day-level and subject-level heterogeneity, the source of emotional variation was mainly the within-day variation, further verifying the necessity for emotional outcome to be monitored in high intensity in a micro environment and such designs as the EMA design can help to achieve this goal.

Table 2.5: Model comparison on data from Adolescent Smoking Study (extended versions)

Parameters	**Model A** $\mathbf{RI}_s\mathbf{+AR(1)}_{ws}$	**Model B** $\mathbf{RI}_s\mathbf{+RI}_d\mathbf{+SAR(1)}_{wd}$	**Proposed Model** $\mathbf{RI}_s\mathbf{+AR(1)}_d\mathbf{+SAR(1)}_{wd}$
$\beta_{Intercept}$	5.214***	5.198***	5.103***
$\beta_{Male(vs.Female)}(L3)$	-0.388**	-0.389**	-0.383**
$\beta_{Grade10}(L3)$	0.103	0.100	0.095
$\beta_{NovSeek}(L3)$	0.212*	0.214*	0.234*
$\beta_{NegMoodReg}(L3)$	-0.806***	-0.803***	-0.792***
$\beta_{AloneBS}(L3)$	0.983**	0.981**	0.961**
$\beta_{AloneWS}(L1)$	0.398***	0.372***	0.373***
$\beta_{Weekend}(L2)$	-0.247***	-0.246***	-0.222***
σ^2_ν	1.745	1.676	1.086
σ^2_ψ		0.583	1.202
σ^2_ϵ	2.912	2.348	2.335
ρ_1	0.438	0.243	0.234
ρ_2			0.737
$-2loglik$	47309	47065	46895
AIC	47331	47089	46921

*** p-value < 0.0001; ** p-value < 0.01; * p-value < 0.05;

Table 2.5 shows the results for extended versions of Model A, B and the proposed model and they yielded similar estimates and significance levels for β as those of the original versions. We still observed the similar decreasing trend of σ^2_ν's from extended Model A, to extended Model B and the extended proposed model. Similarly for the extended models, after including the autocorrelation among random day effects, the subject-level variance decreased by a great amount and it means the level-2 AR(1) explained a large part of the between-subject variance (level-3).

However, including the time-stamped information didn't help to improve the fit to the data as the AICs of the extended versions were greater than AICs of the original versions. As we have shown in the simulation study, the extended version did not necessarily fit the data better than the original version did, especially when the assumption of serial AR(1) errors was violated. Thus for these data, the original version was preferred over the extended model.

2.7 Discussion

In this article, we have proposed a three-level linear model with AR(1) day-level random effects and AR(1) (or serial AR(1)) assessment-level random errors, which provides an interesting idea of the potential correlation hierarchy among assessments. The model also enables researchers to identify sources of variation as well as the (serial) autocorrelation at different levels. The variation in the outcome variable *NEGAFF* in the real EMA study mainly came from the within-day (level-1) variance and the betwee-day variance (level-2), which supports the idea that psychological outcome such as mood should be monitored in a more micro environment. Beyond this, our proposed model not only works for psychological outcomes but also works for other multilevel intensive repeated (irregularly spaced) measures that displayed intensive variation pattern e.g., stock market price.

The time-stamp information has granted more flexibility to the random effect distribution in two aspects. First, it allows the correlation intensity between two within-day assessments to depend on the real longitudinal distance or the length of the interval. Second, because the dimension of the within-day error distribution completely depends on the number of non-missing time-stamps, subjects are allowed to have flexible dimension when missing data are presented. With real time information, we don't even have to know the total number of prompted assessments. This extended model can be a more proper option to handle unbalanced data and especially those with flexible number of completed assessments, e.g., the self-initiated event-contingent responses in EMA study. For those responses, researchers have no information about how many responses will be initiated by the subject.

For the real data, the original version and the extended version of the proposed model had similar estimates for fixed effects but the time-stamps didn't help to improve the fit. The assumption for how we utilized the time-stamps is different for the original and the extended version models. For the original version, time is treated as ordinal/categorical variable; but for the extended version, time is assumed to be continuously linear. But fortunately, according to the simulation study, using smaller -2 Log Likelihood or AIC for model selection is still reliable, which guarantees the correctness for specifying correlation structure.

In the analyses we only discussed models with only the random intercepts but it can also be extended to include random covariate effects if the data contained sufficient information for estimation. We only used the AR(1) structure for the correlated random effects because AR(1) is one of the most excessively used structure in time-series analysis. Under our modelling framework, the choice of correlation structure can be flexible. For example, we can substitute the AR(1) structure by Toeplitz structure as well as autoregressive-moving-average (ARMA) structure which may better depict the reality in EMA data collection. For serial autocorrelation, the function for computing correlation can also be flexibly changed. We assumed the correlation matrix to be the serial power structure of longitudinal distance but we can also assume a Gaussian correlation structure or other useful structures.

For estimation, although the pseudo likelihood approach is much more efficient than the traditional maximum likelihood approach, it has certain limitations as we have discussed in Section 2.4.2. For GLMM, when the outcome doesn't follow a normal distribution, the approximation based on linearization makes the likelihood-based model comparison inapplicable. Therefore, in GLMM setting, traditional integral evaluation using Gaussian Quadrature should be involved in order to obtain the true likelihood value. As an alternative method for estimation, the conventional MMLE with Gaussian Quadrature method takes much longer than the pseudo likelihood approach when large amount of random effects are presented. In the supplemental material, we present the coding scripts of pseudo-likelihood method along with conventional MMLE in SAS syntax using procedure GLIMMIX and NLMIXED for both the original version and the extended version of our model. Harring and Blozis [22] also provided another idea to efficiently estimate the original version for our proposed model using PROC NLMIXED but it was less applicable for the extended version. To expedite the computation, in our programming script for PROC NLMIXED, instead of the Gaussian Quadrature method, we used the first-order method [50]. Future works for improvement can be from aspects such as developing new efficient estimation method for fitting the mixed effect model with complex hierarchical random effects and non-normal responses.

CHAPTER 3

A SHARED-PARAMETER LOCATION-SCALE MIXED MODEL TO LINK THE RESPONSIVITY IN SELF-INITIATED EVENT REPORTS AND THE EVENT-CONTINGENT ECOLOGICAL MOMENTARY ASSESSMENTS

3.1 Introduction

The Ecological Momentary Assessment (EMA) design [51, 26] is useful to explore how participants' psychological states, e.g. mood, and behaviors interact with lasting as well as momentary environmental factors [41]. Subjects in EMA studies complete the real-time assessments on their emotional states or behaviors as well as the surrounding environment [26] via hand-held devices, which helps to reduce recall bias [55, 6] and assess subjects intensively within a day or across days. As a result of this, subjects in EMA studies can have more than 30 within-subject assessments in just a week of study. For EMA studies with mood as the main research interest, such intensive within-subject data contain sufficient information to infer mood change following episodes of events in terms of subject's mood mean as well as mood variability.

Many EMA studies involve two mechanisms to collect assessment data. For the first mechanism, assessments are randomly prompted to participants. While many publications focus on statistical modeling of these prompted assessments [54, 14, 31, 28], the current article focuses on the second mechanism where participants self-initiate an assessment at the moment of the event of interest or behavioral lapse, e.g., smoking. Once participants initiate the survey, they are asked to report on their psychological state, behaviors, and the surrounding environment as well as the social context contingent to the occurring event. Even with effective booster strategies, participants can become less compliant to self-initiated reports during the study. Also, the declining responsivity can be related to the primary longitudinal outcome of interest, e.g., the mood. For example, participants' compliance to self-initiated event reports can fluctuate with their mood level (e.g., subjects do not initiate a report when they feel bad). As a result, the observed self-initiated event-contingent EMA assessments for mood can be subject to bias. We elaborate on this issue with a motivating example where participants' nonresponses can be related to the unobserved outcome in terms of its mean or variance level. The issues described relate to the 'missing not at random' (MNAR) or 'non-ignorable missing' mechanism [33]. MNAR is often thought to be possible for many types of self-report data, regardless of whether the assessments are prompted or self-initiated. However, an extra challenge present in self-initiated assessments is that the nonresponses of such assessments are usually unknown. Because nonresponses are unknown, most of the conventional methods such as imputation [48] or inverse probability weighting (IPW) [49, 34] for missingness are generally inapplicable. However, Kovalchik et al. proposed a scaled version of IPW to correct bias in self-initiated assessments [28] by borrowing information from the responsivity of random prompts. The key assumption of this method is that data are "Missing At Random" (MAR)[33] so that the missing event reports are related to known contextual factors as well as the observed outcome data but not the unobserved outcome.

As a result of these challenges, previous publications have focused almost exclusively on MNAR approaches for prompted assessments rather than for self-initiated assessments. Some methods for the MNAR scenario include the shared parameter mixed model (SPMM) [12, 13] which jointly models the missingness pattern and the primary outcome of interest with models that share the same random effects. In such models, the missingness pattern can be connected to the unobserved outcome through the posterior distribution of the random effects [56, 65, 19]. Also, given the shared random effect, the primary outcome and the missingness pattern are conditionally independent so that the conditional likelihood function can be conveniently constructed. Specifically for EMA data, Cursio et al. have proposed a two-level shared-parameter mixed model with common random subject intercepts in both the mood outcome model and the missingness model [14]. In other words, a subject's nonresponse probability is assumed to be associated with the mean of this subject's mood level. As mood variance is also informative, Lin et al. went further and proposed a shared-parameter location-scale mixed model to link the missingness process to the mean as well as the variance function of the mood outcome via the shared subject-level random intercepts [31].

Besides linking of the random intercepts, these shared-parameter models also allow the possibility of linking the assessment missingness to additional subject-level latent effects. Especially for a longitudinal study, the change or the growth pattern of the outcome over time of each individual can also be another characteristic of interest. Thus, a special case of the multilevel mixed model is the individual growth model [53] which allows subjects to have different slopes in their outcome trajectories over time. The growth model can easily be fit into the shared-parameter framework so that one can explore the potential linkage between responsivity or missingness of assessments with a subject's overall or initial mood level (random intercept) as well as subject's trajectory of mood over time (random slope).

For self-initiated assessments, the only information often obtained to quantify responsivity is the count of responses per day. Replacing the binary responsivity/missingness indicator with such summary measures can be more efficient for estimation [47]. Therefore, we will propose a shared parameter location-scale mixed model to relate subject's mood mean, as well as the mood variance, to the count of observed responses with shared random subject-level effects. The proposed model will be validated by model comparison via a simulation study. Then we will elaborate how our proposed model reveals the potential association between the responsivity of self-initiated reports, and subject's change in mood as well as mood variability. With the study design described in the motivating example [39], we are able to obtain information not only for the responsivity to random prompts but also the extent of responsivity in self-initiated event contingent EMA data (Section 3.2).

3.2 Motivating example

The Dual-Use Study [39] is an EMA study using real-time reporting to explore participants' psychological and behavioral patterns regarding the events of interest, the combustible tobacco use (CIG) and electronic cigarette (ECIG) use. As in most EMA studies, events (of CIG and ECIG use) were obtained via self-initiated reports from the subjects. However, also in this study, events were sometimes converted from the random prompts. In this case, upon being prompted, participants would first be queried if they were experiencing a CIG/ECIG event at the moment, and if so, then the device converted to provide event-related questions. Ideally, if participants were perfectly compliant in self-initiating the reports of every CIG/ECIG event, then none of the events would be captured by random prompts. However, this clearly varied across subjects, and these converted random to event prompts provide some information about missing self-initiated reports that is often not obtained in most of other EMA studies. Therefore, the number of random converted reports can be thought to represent the extent of responsivity/missingness in self-initiated reports.

Many other EMA studies for smoking, and other behaviors, assume these two data collection mechanisms are independent and used in different ways. In contrast, the converted reports in the Dual-Use study provide some information for the self-initiated event-contingent responses that other EMA studies would have missed. Specifically, we will apply our proposed model to the Dual-Use data to examine potential associations between subject's positive mood and the extent of missingness in self-initiated reports.

3.3 The shared parameter approach

Motivated by the Dual-Use study, we propose a shared parameter location-scale mixed model for the random converted event reports. We model the primary outcome with a shared parameter location-scale mixed model [25, 23] as in Lin's model [31]. The main extension of our proposed model from Lin's model is that, instead of modelling the binary missingness indicator of each prompted assessment for responsivity, we model the counts of observed converted responses. Another difference is that, besides including the linkage between subject-level random intercepts in the responsivity model and the primary outcome model, we will also consider the linkage among the random subject time effects in both models.

3.3.1 Primary outcome model

Extending the linear mixed model, Hedeker et al. [25, 23] proposed a location-scale mixed model to additionally model the variances of the mixed model and to include a random subject effect for the within-subject variance. The collected intensive longitudinal data in EMA studies often provides sufficient information for valid estimation of these additional parameters, which may not be possible with data from more standard longitudinal studies. Relaxing the homogeneity assumptions in conventional linear mixed models, subjects are allowed to have different within-subject means (location) as well as different within-subject variances (scale) in location-scale modelling.

In the following notation, vectors and matrices are highlighted in bold. First, data used in the proposed model are three-level. Let Y_{ijk} be the primary longitudinal outcome, e.g., the positive mood level recorded at the k^{th} assessment of subject i at day j. $\boldsymbol{\beta}$ is the $(p+1)$-dimension fixed effects vector consisting of the population intercept and p covariate effects. Particularly for the growth model setting, time effects may be characterized by polynomials or other comparisons across time and are included in $\boldsymbol{\beta}$. $\boldsymbol{X}$ is the design matrix including subject-level, day-level and assessment-level covariates. The random subject effects $\boldsymbol{\nu_i} = (\nu_{i0}, \nu_{i1}, \ldots, \nu_{iq})^T$ is a $(q+1)$-dimension vector in the mean function of the primary outcome and are called the random location effects (Equation (3.1)). Such a vector consists of the random subject intercept as well as random effects for q selected within-subject (level-1 or level-2) covariates, representing how each subject deviates from the population effects in $\boldsymbol{\beta}$. Specifically, in the linear growth model setting, $\boldsymbol{\nu_i} = (\nu_{i,int}, \nu_{i,time})^T$ correspond to random intercept and random linear change (increment) for subject i. Z is the level-1 or level-2 known covariate matrix where the elements in first column of $\boldsymbol{Z}$ are all 1's corresponding to the subject-level random intercepts. It is assumed that $\boldsymbol{\nu_i} \sim \mathcal{MVN}(0, \boldsymbol{\Sigma_\nu})$. Non-zero covariance among the random location effects is allowed. But across subjects, ν_i are independent of one another.

$$y_{ijk} = \boldsymbol{X_{ijk}}^T\boldsymbol{\beta} + \boldsymbol{Z_{ijk}}^T\boldsymbol{\nu_i} + \epsilon_{ijk} \tag{3.1}$$

In traditional linear mixed models, the random errors ϵ_{ijk} are assumed to be identically distributed. However, as has been observed in many studies, some participants' mood fluctuated more erratically across time than others. Thus, in Equation (3.2), the error variance is formulated by a ln-linear model and $\boldsymbol{\alpha}$ is the vector of fixed effects that may impact the variability of level-1 errors. For example, age can be a factor related to the error variability that younger people may have higher mood variability than elderly people or in other words, elderly people may have higher mood stability.

Besides the possible fixed effects $\boldsymbol{\alpha}$, a latent variable ω_i is included, which summarizes other unobserved subject traits independent from the observed variables and is also called the random scale effect. The random scale effect may contribute to the heterogeneity of level-1 errors. It is assumed that $\omega_i \sim \mathcal{N}(0, \sigma^2_\omega)$ so that the error variance is assumed to follow a ln-normal distribution.

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W_{ijk}}^T \boldsymbol{\alpha} + \omega_i) \tag{3.2}$$

The location random effects and scale random effects are allowed to be dependent, see Equation (3.3). The subject mood mean can be associated with the mood variability by assuming non-zero covariance $\boldsymbol{\Sigma_{\nu\omega}}$ for $\boldsymbol{\nu_i}$ and ω_i. In prior psychological studies of positive mood, a negative correlation between the mean of mood and the variance was often observed [20], such that subjects with higher positive mood were more consistent in their responses.

$$\begin{pmatrix} \boldsymbol{\nu_i} \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_\nu} & \boldsymbol{\Sigma_{\nu\omega}} \\ \boldsymbol{\Sigma_{\nu\omega}}^T & \sigma^2_\omega \end{pmatrix}\right] \tag{3.3}$$

The location-scale mixed model allows us to explore the subject-level mood mean and the mood variability as well as the association between these two quantities.

3.3.2 Responsivity model

The total count of (converted) responses of subject i on day j, C_{ij}, reflects the degree of responsivity to the self-initiated assessments and is another outcome for modelling.

Specifically, we model the counts of (converted) responses (C_{ij}) and assume C_{ij} follows a Poisson distribution with a positive mean formulated by an exponential function, in which there is a ln-linear equation parameterized by the covariates and the random subject effect λ_i (Equation (3.4)(3.5)). An offset term t_{ij} can also be added just as in other Poisson regression models to adjust for exposure level.

$$C_{ij} \sim \mathcal{POI}(\mu_{ij}) \tag{3.4}$$

$$\ln(\frac{\mu_{ij}}{t_{ij}}) = \boldsymbol{L_{ij}}^T \boldsymbol{\tau} + \lambda_i \tag{3.5}$$

3.3.3 Shared random effects

In the sense of parameter sharing, the random subject effects $\boldsymbol{\nu_i}$, ω_i, and λ_i in Equations (3.1), (3.2), and (3.5) are assumed to be correlated by a set of linkage linear equations (Equation (3.6)). To simplify, λ_i in Equation (3.5)) is linearly associated with $\boldsymbol{\nu_i}$ and ω_i respectively and the intensity of association is quantified by $\boldsymbol{\gamma}$ and δ. Again, these random subject effects, $\boldsymbol{\nu_i}$, ω_i, and λ_i summarize some latent subject-specific features that may affect both the mood mean, mood variability, and also the likelihood of responsivity. In our model, rather than to linearly combine these two linkage components in a single equation, we will use two separate sets of equations in order to better observe the effects regarding mood location and mood scale independently.

The sign of the linkage parameters $\boldsymbol{\gamma}$ and δ indicates positive/negative association between the responsivity and subject's mood location and scale. As we have discussed in Section 3.3.3, if we assume $\boldsymbol{\gamma} = (\gamma_{int}, \gamma_{time})$, a positive linkage ($\gamma_{int}$) between λ_i and $\nu_{i,int}$ indicates that higher subject mood levels are associated with more initiated/converted reports and vice versa; while a positive linkage (γ_{time}) between λ_i and $\nu_{i,time}$ indicates that the amount of subject's mood increase over time is associated with the subject's responsivity. Therefore, besides subject's average mood level, the amount of change can contribute to higher or lower responsivity of the self-initiated reports.

As we showed in Equation (3.3) the location and scale effects are allowed to be correlated. In this specification, $\boldsymbol{\nu_i}$ and ω_i are correlated inherently since they shared the same λ_i. $\boldsymbol{\eta_{1i}}$ and η_{2i} are additional parameters following independent normal distributions across subjects with mean 0 and variance $\boldsymbol{\Sigma_{\eta_1}}$ and $\sigma^2_{\eta_2}$, respectively, to absorb the extra variability unexplained by the variation in λ_i. Note that $\boldsymbol{\Sigma_{\eta_1}}$ is a diagonal matrix.

$$\begin{cases} \boldsymbol{\nu_i} = \lambda_i \boldsymbol{\gamma} + \boldsymbol{\eta_{1i}} \\ \omega_i = \lambda_i \delta + \eta_{2i} \end{cases} \tag{3.6}$$

3.4 Estimation

3.4.1 Conditional likelihood functions

According to Equation (3.1) and (3.2), given $\boldsymbol{\nu_i}, \omega_i$, the conditional likelihood of the random errors $\boldsymbol{\epsilon_{ij}} = (\epsilon_{ij1}, \epsilon_{ij2}, \ldots, \epsilon_{ijN_{ij}})^T$ for day j in subject i is,

$$f(\boldsymbol{\epsilon_{ij}}|\boldsymbol{\nu_i}, \omega_i) = \frac{1}{\sqrt{|2\pi \boldsymbol{\Sigma_{\epsilon_{ij}}}|}} \exp(-\frac{1}{2}\boldsymbol{\epsilon^T_{ij}} \boldsymbol{\Sigma^{-1}_{\epsilon_{ij}}} \boldsymbol{\epsilon_{ij}}) \tag{3.7}$$

$\boldsymbol{\Sigma_{\epsilon_{ij}}}$ is a diagonal matrix of size $N_{ij} \times N_{ij}$. Each its diagonal element $\sigma^2_{\epsilon_{ijk}}$ takes the form as below,

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W^T_{ijk}} \boldsymbol{\alpha} + \omega_i) \tag{3.8}$$

Then we substitute $\boldsymbol{\nu_i}, \omega_i$ with λ_i and $\boldsymbol{\eta_{1i}}$ and η_{2i} according to Equation (3.6),

$$\epsilon_{ijk} = Y_{ijk} - \boldsymbol{X_{ijk}^T}\boldsymbol{\beta} - \boldsymbol{Z_{ijk}^T}(\lambda_i\boldsymbol{\gamma} + \boldsymbol{\eta_{1i}}) \tag{3.9}$$

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W_{ijk}^T}\boldsymbol{\alpha} + (\lambda_i\delta + \eta_{2i})) \tag{3.10}$$

so that the conditional likelihood $\boldsymbol{\epsilon_{ij}}$ of is,

$$f(\boldsymbol{\epsilon_{ij}}|\lambda_i, \boldsymbol{\eta_{1i}}, \eta_{2i}) = \prod_k \frac{1}{\sqrt{2\pi\sigma^2_{\epsilon_{ijk}}}} \exp(-\frac{\epsilon^2_{ijk}}{2\sigma^2_{\epsilon_{ijk}}}) \tag{3.11}$$

For the count outcome, the likelihood function conditioning on λ_i is,

$$f(C_{ij}|\lambda_i) = \frac{\mu_{ij}^{C_{ij}} \exp(-\mu_{ij})}{C_{ij}!} \tag{3.12}$$

,

and $\mu_{ij} = exp(\boldsymbol{L_{ij}^T}\boldsymbol{\tau} + \lambda_i)$ according to Equation (3.5) and the distribution of λ_i is,

$$g(\lambda_i) = \frac{1}{\sqrt{2\pi\sigma^2_\lambda}} \exp(-\frac{\lambda_i^2}{2\sigma^2_\lambda}) \tag{3.13}$$

The joint distribution of $\boldsymbol{\eta_{1i}}$ and η_{2i} is,

$$g(\boldsymbol{\eta_{1i}}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma_{\eta_1}}|}} \exp(-\frac{1}{2}\boldsymbol{\eta_{1i}^T}\boldsymbol{\Sigma_{\eta_1}^{-1}}\boldsymbol{\eta_{1i}}) \tag{3.14}$$

$$g(\eta_{2i}) = \frac{1}{\sqrt{2\pi\sigma^2_{\eta_2}}} \exp(-\frac{\eta^2_{2i}}{2\sigma^2_{\eta_2}}) \tag{3.15}$$

And $\boldsymbol{\epsilon_{ij}}$ and C_{ij} conditioning on $\lambda_i, \boldsymbol{\eta_{1i}}, \eta_{2i}$ are independent. Thus, the conditional joint likelihood function for both outcomes is

$$f(\boldsymbol{\epsilon_{ij}}, C_{ij}|\lambda_i, \boldsymbol{\eta_{1i}}, \eta_{2i}) = f(C_{ij}|\lambda_i) f(\boldsymbol{\epsilon_{ij}}|\lambda_i, \boldsymbol{\eta_{1i}}, \eta_{2i}) \tag{3.16}$$

The marginal joint likelihood for ϵ and C is,

$$\mathcal{L}(\boldsymbol{\epsilon, C}) = \int \prod_i \{ \prod_j f(C_{ij}|\lambda_i) f(\boldsymbol{\epsilon_{ij}}|\lambda_i, \boldsymbol{\eta_{1i}}, \eta_{2i}) \} g(\lambda_i) g(\boldsymbol{\eta_{1i}}) g(\eta_{2i}) d\lambda_i d\boldsymbol{\eta_{1i}} d\eta_{2i} \tag{3.17}$$

3.4.2 Optimization

From Section 3.4.1, integrations over the random effects needs to be performed to construct the marginal likelihood function. For generalized linear models this integral usually has no closed form. Thus, numerical integral evaluation techniques such as Adaptive Gaussian Quadrature [44, 30] are needed. The Adaptive Gaussian Quadrature technique approximates this integral by a weighted linear sum of the conditional likelihood evaluated at multiple chosen quadrature points and the corresponding weights are calculated based on the shape of random effect distributions. Additionally, we used the Newton-Raphson algorithm with ridging (NRRIDG) to optimize the approximated marginal log likelihood function. For this, estimation can be achieved using Procedure NLMIXED in SAS 14.3 (SAS Inc., Cary, NC).

The appendix provides some sample code illustrating this approach.

3.5 Simulation

3.5.1 Models for comparison

We compared our proposed model to two more conventional models. In this simulation study, without loss of generality, $\boldsymbol{\nu_i}$ in the mean function includes only the random intercept $\nu_{i,int}$ in this simulation study and therefore the vectors $\boldsymbol{\nu_i}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta_{1i}}$ are reduced to scalars, i.e., $\nu_{i,int}$, γ_{int} and $\eta_{1i,int}$, and the $\boldsymbol{Z}$ matrix is reduced to an all-ones column vector.

One naïve model is the linear mixed model with heterogeneous variance and only subject-level random location effect (HVMM), where the number of reports is not related to a subject's mood. We note that there is no random scale effect and the corresponding linkage. In this model, $\boldsymbol{\gamma}$ was set to equal zero so that the responsivity model was assumed to be independent of the mood model. In the second model, we allowed some association between a subject's mood and the number of reports but only via the mood location. Still, compared to our proposed model, this model doesn't have random scale effect. For this approach, we estimated $\boldsymbol{\gamma}$ in the mean function but not the scale linkage δ (Shared Parameter Mixed Model-Location, SPMM-L).

Heterogeneous Variance Mixed Model (HVMM):

$$y_{ijk} = \boldsymbol{X_{ijk}^T\beta} + \boldsymbol{Z_{ijk}^T\nu_i} + \epsilon_{ijk} \tag{3.18}$$

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W_{ijk}^T\alpha}) \tag{3.19}$$

$$ln(\frac{\mu_{ij}}{t_{ij}}) = \boldsymbol{L_{ij}^{T}\tau} + \lambda_i \tag{3.20}$$

Shared Parameter Mixed Model-Location (SPMM-L):

$$y_{ijk} = \boldsymbol{X_{ijk}^{T}\beta} + \boldsymbol{Z_{ijk}^{T}}(\lambda_i \boldsymbol{\gamma} + \boldsymbol{\eta_{1i}}) + \epsilon_{ijk} \tag{3.21}$$

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W_{ijk}^{T}\alpha}) \tag{3.22}$$

$$\ln(\frac{\mu_{ij}}{t_{ij}}) = \boldsymbol{L_{ij}^{T}\tau} + \lambda_i \tag{3.23}$$

Shared Parameter Location/Scale Mixed Model (SPMM-LS)

$$y_{ijk} = \boldsymbol{X_{ijk}^{T}\beta} + \boldsymbol{Z_{ijk}^{T}}(\lambda_i \boldsymbol{\gamma} + \boldsymbol{\eta_{1i}}) + \epsilon_{ijk} \tag{3.24}$$

$$\sigma^2_{\epsilon_{ijk}} = \exp(\boldsymbol{W_{ijk}^{T}\alpha} + (\lambda_i \delta + \eta_{2i})) \tag{3.25}$$

$$\ln(\frac{\mu_{ij}}{t_{ij}}) = \boldsymbol{L_{ij}^{T}\tau} + \lambda_i \tag{3.26}$$

3.5.2 Simulation setting

We validated the proposed model and compared three mentioned candidate models on 100 datasets simulated respectively from two scenarios. In the first scenario, responsivity was not related to the primary outcome in terms of location or scale; this scenario was in the similar sense to MAR scenario. Thus, we set the true values of $\boldsymbol{\gamma}$ and δ to zero, which means no relationship between the primary longitudinal outcome and the number of responses. In the second scenario, we set non-zero values for $\boldsymbol{\gamma}$ and δ and allowed both a subject's mean and variance to be associated with the number of responses, which mimicked an MNAR process. The 100 simulated datasets had the same three-level data structure as the real data, that is, there were 100 subjects in each dataset and each subject had data from 7 days. The count of (converted) responses of each day was generated from a Poisson distribution. We simulated covariates for each of the three levels, respectively, for the primary outcome model. For the responsivity model (level 2), only the subject-level (level-3) and day-level (level-2) covariates were included.

We compared the results from all three models (the more conventional HVMM and SPMM-L, and the proposed SPMM-LS) in terms of the average of the parameter estimates and the coverage rate for each parameter. The coverage rate was computed as the proportion of solutions in which the 95% confidence interval contained the true value.

3.5.3 Simulation Results

Under the first scenario, with completely no linkage between the primary outcome and the responsivity (Table 3.1), all three models resulted in reasonable bias and correct coverage on the main parameters. However, for the second scenario (Table 3.2) parametrized by our proposed model SPMM-LS, both the naïve HVMM model and SPMM-L had low coverage

Table 3.1: Model comparison over 100 simulated datasets from the first scenario (HVMM)

Parameters	True	HVMM		SPMM-L		SPMM-LS	
		Mean	CR	Mean	CR	Mean	CR
Mood Response Model (LS Mixed Model)							
β_0	5	5.02	0.92	5.02	0.91	5.02	0.91
β_1	1	0.99	0.95	0.99	0.96	0.99	0.96
β_2	-2	-2.00	0.94	-2.00	0.94	-2.00	0.95
β_3	1	1.00	0.93	1.00	0.93	1.00	0.93
α_0	1	1.00	0.94	1.00	0.94	1.00	0.94
α_1	-0.5	-0.50	0.96	-0.50	0.96	-0.50	0.96
α_2	0.1	0.10	0.97	0.10	0.97	0.10	0.98
α_3	-0.3	-0.30	0.96	-0.30	0.96	-0.30	0.96
Linkage							
γ	0			0.01	0.93	0.01	0.93
δ	0					0.00	0.93
Count/Rate Model (Poisson Mixed Model)							
τ_0	0.7	0.71	0.99	0.71	0.99	0.71	0.99
τ_1	0.5	0.49	0.97	0.49	0.97	0.49	0.97
τ_2	-0.5	-0.50	0.96	-0.50	0.96	-0.50	0.96
$ln(\sigma_\lambda^2)$	0.5	0.50	0.94	0.50	0.94	0.50	0.94

True: true value;
CR: coverage rate;

on α the error variance parameters. And for the SPMM-L, removing the scale random effect and the linkage on scale affected coverage on the location linkage. In contrast, our proposed model successfully recovered all of the model parameters with correct coverage and acceptable bias.

Although the linkage specification across three models were different, the estimates in the responsivity model component were similar, no matter in which scenario.

3.6 A Shared-Parameter Location-Scale Mixed Model for Dual-Use Study

As we mentioned in Section 3.2, in Dual-Use Study, the event reports converted from random prompts represented an observed part of the missing self-initiated reports. We focused on

Table 3.2: Model comparison over 100 simulated datasets from the second scenario (SPMM-LS)

Parameters	**True**	**HVMM**		**SPMM-L**		**SPMM-LS**	
		Mean	**CR**	**Mean**	**CR**	**Mean**	**CR**
Mood Response Model (LS Mixed Model)							
β_0	5	5.06	0.91	4.95	0.95	4.98	0.95
β_1	1	0.98	0.93	1.01	0.97	1.00	0.97
β_2	-2	-2.00	1.00	-2.00	1.00	-2.00	0.98
β_3	1	1.00	0.97	1.00	0.97	1.00	0.95
α_0	1	0.30	0.04	0.30	0.04	1.00	0.95
α_1	-0.5	-0.48	0.11	-0.48	0.11	-0.50	0.94
α_2	0.1	0.08	0.19	0.08	0.20	0.10	0.95
α_3	-0.3	-0.31	0.46	-0.31	0.46	-0.30	0.96
Linkage							
γ	1			1.02	0.87	1.00	0.95
δ	-1					-0.99	0.91
Count/Rate Model (Poisson Mixed Model)							
τ_0	0.7	0.71	0.99	0.69	0.98	0.71	0.99
τ_1	0.5	0.49	0.98	0.49	0.98	0.49	0.97
τ_2	-0.5	-0.5	0.95	-0.5	0.95	-0.5	0.97
$ln(\sigma^2_\lambda)$	0.5	0.47	0.96	0.69	0.98	0.47	0.96

True: true value;
CR: coverage rate;

modelling the number of this type of responses because these converted prompts reflect the extent of responsivity of the self-initiated reports.

3.6.1 The Data of Dual-Use Study

There were 287 subjects in the Dual-Use Study [39] and 57.84% of the sample were males and 40.77% were females, the average age of the sample was 35.42 years. Subjects in this study were instructed to self-initiate and report on combustible tobacco use (CIG) and non-combustible tobacco use (ECIG) over a seven-day study period. In addition, they also received random prompts in which they were queried about whether they were smoking CIG/ECIG at that moment and if so, the random prompts were converted into event reports.

Other than subject characteristics such as age and gender, variables associated with subject's smoking habit were recorded as well, such as the Nicotine Dependence Syndrome Scale (CigNDSS) score and 30-day smoking rate (CigRate) measured prior to baseline. All these subject-level continuous variables were centeringed in terms of sample means for better interpretability. The time variable, denoted as studyday, ranged from 1 to 7. For any given day, participants had from 1-13 self-initiated events (SI events) and 1-7 events which were converted prompts (RC events).

We jointly modeled two outcomes: the first was a continuous outcome (positive mood before CIG smoking) assumed to follow a normal distribution parametrized by the location-scale mixed model; the second was the number of converted event reports of CIG smoking following a Poisson distribution formulated by a Poisson mixed effect model with an offset. The offset variable was defined as the number of times per day that the subject used the hand-held device to either self-initiate reports or complete prompted questions regarding Cig or ECIG use.

3.6.2 Results

The candidate models to be compared for the Dual-Use data are similar to those in the simulation study. However, in terms of the linkage, we not only linked the random intercept in each model component, we also specified linkage between the random intercept in the responsivity model and the random time effect(s) in the mood location function. For each model, we obtained the estimates along with the 95% confidence intervals for each of the model parameters. Results of model comparison are summarized in Table 3.3. Lower AIC and -2 log likelihood (-2LL) values are associated with better performance of the model and imply that the data contain sufficient information to model the mood location (random intercept and time effect(s)), mood scale (random mood variability) and their association on the number of converted events (responsivity).

Results are shown in Table 3.3, with significant covariate effects highlighted in bold. All of the three candidate models include random intercept and random linear time effects. However, HVMM doesn't allow any linkage to responsivity or random scale effect, SPMM-L only allows for the location linkage but still there is no random scale effect, and the proposed SPMM-LS model has both random location and scale effects and allows linkage in terms of both location and scale.

As one can see, model SPMM-L is slightly better than HVMM which indicates that a subject's mood level/mood change was associated with nonresponses in the self-initiated assessments. However, the biggest improvement comes from including the random scale effect and its linkage. Our proposed model is shown to have the best fit to the data with the lowest AIC value.

For the fixed covariates effects β, the three models yielded similar estimates and 95% confidence intervals. Older participants and females displayed more positive mood on average. CigNDSS (smoking dependency) had a significant negative association with positive mood while the effect of CigRate was not significant. The linear time effect (studyday) was negative and significant across all three models and indicated that mood became less positive over the days of the study.

For covariates effects α in the within-subject variance function, the two naïve models yielded similar estimates and confidence intervals. Across all three models, older subjects had more stable mood than younger subjects, and females had higher mood variability than males. While CigNDSS had a significant positive association with higher mood variability, higher CigRate at baseline was significantly associated with lower variability in the two naïve approaches. However, including the random scale effect and the linkage of the proposed model, CigRate became insignificant. In terms of studyday, subject's mood variability decreased over days as the linear time effect is negative and significant.

Three models produced similar estimates and confidence intervals for the responsivity component where age, gender, CigRate and the linear day effect were significant. Older subjects had more converted responses than younger subjects, i.e., younger subjects may be more compliant to self-initiate event reports. Male subjects had more converted responses than female subjects, indicating that males may be less compliant. Positive effects for CigRate and CigNDSS indicate that higher CigNDSS and CigRate were related to more converted events or less compliance. The positive and significant linear day effect implies that participants became less compliant to initiate the event reports, resulting in more events being captured by the random prompts.

The linkage parameters summarize the latent relationship between subject's mood location, mood scale, and responsivity at the subject level. Our proposed model (SPMM-LS) was able to reveal significant associations between 1) the mood scale and responsivity and 2) between linear mood change and responsivity. The scale linkage was significant and positive, which indicates that higher mood variability was associated with lower compliance as more events were obtained by conversion instead of being reported by the subject. The linkage of the linear mood change and responsivity was also positive and significant. The greater degree that a person's mood increased over time, the more events were obtained by conversion.

3.7 Conclusion

In this article we further extend the location-scale mixed model of Hedeker et al. [25, 23] under the framework of a shared-parameter model to jointly model the multilevel intensive

Table 3.3: Model Comparison on Dual-Use Data: HVMM, SPMM-Location and SPMM-Location/Scale of Linear Growth Model Version

Param	**HVMM**		**SPMM-L**		**SPMM-LS**	
	Est.	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**
Mood Response Model (LS Mixed Model)						
β_{int}	**6.41**	**(6.12, 6.71)**	**6.43**	**(6.14, 6.73)**	**6.40**	**(6.11, 6.69)**
β_{age}	**0.03**	**(0.02, 0.05)**	**0.03**	**(0.02, 0.05)**	**0.04**	**(0.02, 0.05)**
β_{female}	-0.19	(-0.64, 0.25)	-0.20	(-0.64, 0.25)	-0.20	(-0.65, 0.24)
$\beta_{CigNDSS}$	**-0.43**	**(-0.77,-0.09)**	**-0.43**	**(-0.77, -0.10)**	**-0.45**	**(-0.78,-0.11)**
$\beta_{CigRate}$	0.02	(-0.01, 0.06)	0.02	(-0.01, 0.06)	0.02	(-0.01, 0.06)
β_{time}	**-0.07**	**(-0.10,-0.03)**	**-0.07**	**(-0.11,-0.04)**	**-0.07**	**(-0.10,-0.04)**
α_{int}	**0.75**	**(0.61, 0.89)**	**0.75**	**(0.62, 0.89)**	**0.57**	**(0.37, 0.77)**
α_{age}	**-0.01**	**(-0.01,0.00)**	**-0.01**	**(-0.01,-0.00)**	**-0.01**	**(-0.02,-0.00)**
α_{female}	**0.23**	**(0.11, 0.35)**	**0.23**	**(0.11, 0.35)**	**0.42**	**(0.18, 0.66)**
$\alpha_{CigNDSS}$	**0.47**	**(0.38, 0.56)**	**0.47**	**(0.39, 0.56)**	**0.43**	**(0.24, 0.61)**
$\alpha_{CigRate}$	**-0.02**	**(-0.02 , -0.01)**	**-0.02**	**(-0.02, -0.01)**	-0.01	(-0.03, 0.01)
α_{time}	**-0.07**	**(-0.10 ,-0.04)**	**-0.07**	**(-0.10,-0.04)**	**-0.12**	**(-0.15,-0.09)**
Linkage						
γ_{int}			0.08	(-0.38, 0.55)	0.03	(-0.42, 0.48)
γ_{time}			0.06	(-0.01, 0.14)	**0.07**	**(0.01, 0.13)**
δ_{int}					**0.30**	**(0.07, 0.53)**
Responsivity						
τ_{int}	**-2.15**	**(-2.28,-2.02)**	**-2.15**	**(-2.28,-2.02)**	**-2.15**	**(-2.29,-2.02)**
τ_{age}	**0.01**	**(0.00, 0.02)**	**0.01**	**(0.00, 0.02)**	**0.01**	**(0.00, 0.02)**
τ_{female}	**-0.20**	**(-0.37,-0.03)**	**-0.20**	**(-0.37,-0.03)**	**-0.20**	**(-0.37,-0.03)**
$\tau_{CigNDSS}$	0.01	(-0.01, 0.03)	0.01	(-0.01, 0.03)	0.01	(-0.01,0.03)
$\tau_{CigRate}$	**0.30**	**(0.17, 0.42)**	**0.30**	**(0.17, 0.42)**	**0.30**	**(0.18,0.43)**
τ_{time}	**0.03**	**(0.02, 0.04)**	**0.03**	**(0.02, 0.04)**	**0.03**	**(0.02,0.04)**
Goodness of Fit						
-2LL	13828		13823		13502	
AIC	13870		13869		13552	

Param: Parameters;
Est.: Estimate;
95% CI: Confidence Interval

self-initiated mood assessments contingent to CIG use and the responsivity of these mood assessments. This model can also be viewed as an extension of a longitudinal individual growth model by including linkage among the latent subject-specific change parameters and the responsivity of the assessments. As was observed in our data analysis, the covariates of time, gender, age and CigRate were allowed to be associated with the converted responses, which represent part of the nonresponses in self-initiated assessments. For each individual, the mood change as well as the variability was also allowed to be associated with their responsivity.

This model provides additional insights in the modeling of intensive longitudinal assessments. First, it enables us to model the variability of the mood, subjects' mood changing pattern, and the responsivity at the same time. In addition, it allows common covariates as well as common latent subject effects of interest to have impact on these three different components so that the inter-correlation network among covariates and outcomes can be revealed.

In particular, with an EMA design such as the design of the Dual-Use study, our proposed model allows us to infer the psychological reasons for the missingness of the self-initiated event reports, which are often ignored in most EMA studies. Without knowing the missingness of each self-initiated report, the responsivity has been quantified as the counts of the RC reports. Although we only presented the results for RC events, the proposed model can also be used for modelling the observed counts of event reports (RC+SI) or other types of targeting events as other types of responsivity measures.

3.8 Discussion

Using shared parameter models to model responsivity may have limitations. As Little [33] showed, the shared parameter framework will be valid if the nonresponse mechanism is 'random-coefficient-dependent MNAR (RCD-MNAR)'. In RCD-MNAR, it assumes the shared random effects contain all information of the missing outcomes so that the conditional independence between the missing data process and the measurement process can be achieved. This RCD-MNAR mechanism is a special case of a more general and realistic scenario, the "outcome-dependent MNAR (OD-MNAR)" and he stated that the two MNAR mechanisms need to be distinguished. However, for OD-MNAR, the shared parameter model may be insufficient and the covariate effect estimates can biased. Gottfredson et al. [21] evaluated the shared parameter model in the growth model setting under these two MNAR scenarios and showed that the shared parameter model estimates are unbiased only when the missing mechanism is RCD-MNAR. Therefore, if the data are missing as OD-MNAR, we may still need to use other bias correction approaches.

The maximum likelihood (ML) approach we used provides a frequentist way for inference as compared to the Bayesian approach. By ML methods, we are able to obtain standard errors and confidence intervals and p-values, which are widely used for obtaining conclusions in data analysis projects. The shortcoming of ML methods is that it includes a time-consuming multi-dimension integration over the random effects when using the Adaptive Gaussian Quadrature for integral evaluation. In fact, the complexity of the integration algorithm grows exponentially as the dimension of random effect grows. In addition, the convergence of the ML solution can be sensitive to the initial values of the optimization when using the NLMIXED Procedure in SAS to fit such location-scale mixed models. Especially when the target model has many more covariates, the convergence may be more difficult to achieve. Based on this, we may need to select only the informative variables to simplify the model prior to the modelling stage for efficient convergence.

In addition, for simplicity, we have only considered a two-level model (within-subject and between-subject) for the mood outcome. If the data contain sufficient structural information, a three-level location-scale model will also be allowed. In addition, this shared parameter approach can be further adapted to a structural equation modeling approach to account for multiple outcomes.

CHAPTER 4

A MIXED EFFECT LOCATION-SCALE MODEL WITH MIXTURE DISTRIBUTED RANDOM SCALE EFFECTS TO ANALYZE (NEAR-) IDENTICAL ENTRIES IN ECOLOGICAL MOMENTARY ASSESSMENTS

4.1 Introduction

EMA assessments (level-1 units) are usually collected intensively for each participant (level-2 units) via self-reporting using portable handheld devices. These highly intensive data provide sufficient information to model the heterogeneity in not only subjects' means (location) but also variances (scale) using Mixed Effect Location-Scale Model (MELS) [23]. However, similar to other types of intensive longitudinal survey data and especially those conveniently collected by handheld devices, EMA data can also be confronted with the issue of (near-) identical within-subject responses. Identical responses can emerge due to: (1) the data are careless responses [38], e.g., subjects get fatigued to the intensive survey so that they provide identical responses; (2) the data are actually valid, e.g., certain subjects do have much more consistent longitudinal outcomes than others. In this case, just naively including these data containing the careless responses may bias the estimation of the Mixed Effect Location-Scale Model. Another action we can take is to remove the data but at the same time the data can be valid data that can contribute to estimation and inference. But no matter what the source of identical response is, the estimation of the MELS can have computational difficulties if the amount of identical responses isn't ignorable. For example, optimization of the Mixed Effect Location-Scale Model will become much more sensitive to the choice of initial values and the optimization may not converge.

Thus, how we handle the identical response may have impact on the estimation, inference as well as the computation. One assumes that some of the subjects may come from another population with outcomes of much higher consistency. To translate this issue into the MELS framework where the random subject effects in the mean function (location random effects) and the variance function (scale random effects) are both modeled, we assume that the scale random effects are more likely to follow a multi-modal distribution rather than a uni-modal zero-centeringed Gaussian distribution. Misspecification of the random effect distribution may bring in substantial biases in the maximum likelihood estimates especially the variability parameter estimates[60]. Type 1 error and also the statistical power can be severely affected by the misspecification of the random effect distribution [32].

There are possible solutions for random effect distribution misspecification. First, one can assume a non-parametric distribution for the random subject effects[42] but Agresti et al. argued that a non-parametric distribution is generally less efficient than a parametric distribution that resembles the true distribution [1] . Another possible approach is to use a semi-parametric approach [9] but model comparisons using log-likelihood based quantities such as deviance and AIC/BIC will not be available. Given such benefits, parametric distributions are still preferred over non-parametric distributions. People applied different parametric random effects distributions in different research contexts. Zhang et al. used a log-gamma distribution to account for different levels of skewness in the observed distributions respectively for random intercept and random slope [66] . Piepho and McCullogh used a parametric family to account for arbitrary distributions of different combinations of kurtosis and skewness[43]. These complicated non-normal distributions are definitely applicable for the peculiar shapes of the random effects distribution but they are less informative for clustering and difficult to interpret.

Among the parametric distributions, the Gaussian mixture distribution is intuitive to account for the irregular distributional shape [18] of random effects and provides a lot of benefits. First, finite mixture of Gaussian distributions is informative to classify the potential subgroups in the population. Second, each Gaussian component has the same functional form so that distributional parameters such as cluster mean and variance are comparable. Third, the estimation method such as Expectation-Maximization (EM) for Gaussian mixtures has been well developed [46]. As a result, Gaussian mixture has been applied to random effects in generalized linear model setting [58]. Xu and Hedeker proposed a mixed model with random subject effects following a mixture of Gaussian distributions [62], which was further used to classify subjects into sub-groups with similar temporal pattern of treatment responses. What warrants the new extension to this model is that, latent clustering may exist in the variability level rather than the mean level, and we assume the mean level across clusters is the same but clusters have different levels of dispersion in the error distribution. Hence, in our proposed model, this idea of Gaussian mixture random effects will be applied to extend MELS by specifying the mixture normal distribution for the random scale effects.

Given these data, we aimed to develop a model to distinguish difference variance patterns among subjects and compute the posterior probabilities of cluster membership. We first validated our proposed model in different simulation scenarios, e.g., different combinations of cluster means and membership proportions. And then we applied the proposed method to a real dataset which contained perfectly identical within-subject responses and compared its fixed effect estimates to those obtained from conventional MELS.

4.2 The Proposed Method

This proposed method will be an extension of the two-level MELS [23] (Equation (4.1)) with subject level location random effects in the mean function and scale random effects in the variability function of the response. Conventionally, the random location effect ν_i and the random scale effect ω_i follow uni-modal zero-mean normal distributions, e.g., $\nu_i \sim \mathbf{N}(0, \sigma_\nu^2)$ and $\omega_i \sim \mathbf{N}(0, \sigma_\omega^2)$.

$$
\begin{aligned}
y_{ij} &= X_{ij}^T \beta + \nu_i + \epsilon_{ij} && (4.1) \\
\sigma_{\epsilon_{ij}}^2 &= \mathbf{exp}(W_{ij}^T \tau + \omega_i) && (4.2)
\end{aligned}
$$

In our proposed model, the random location effect ν_i still follows a single normal distribution with mean zero. But for scale random effect ω_i we now relax the zero-mean normality assumption and assume the scale random effect may follow a mixture of normal distributions [37] so that proposed model is also called as MELS with mixture distributed random scale effects (MELS-MS). Without loss of generality, we will specify the two-cluster version of Gaussian mixture distribution to be our scale effects distribution (Equation (4.3)).

$$
\omega_i \sim \pi \mathcal{N}(\mu_1, \sigma_{\omega_1}^2) + (1 - \pi) \mathcal{N}(\mu_2, \sigma_{\omega_2}^2) \tag{4.3}
$$

The parameter π to be estimated is the marginal mixture probability that a subject have reasonable within-subject variance. We note that $P(C_i = k) = \pi_k$, the probability of being in the cluster k. For the two-cluster version, while subjects in the first cluster display reasonable variability and μ_1 may be close to zero, the estimate of μ_2 can be a very negative value (like -10) which is able to scale the error variance in Equation (4.1) to almost 0. However, estimating the mixture proportion π_k can be computationally challenging.

4.3 Estimation and Inference

For estimation, we incorporated Maximize-A-Posteriori (MAP) algorithm [5] into the Expectation-Maximization (EM) algorithm framework [15] so that we can estimate the posterior probability of clustering membership for each subject. MAP algorithm optimizes the joint likelihood of both the random effects and the random errors [5] over fixed effects and random effects iteratively and therefore bypasses the time-consuming integral approximation approach in each iteration. Next, given the unbiased MAP estimates and random effects, we implemented a few steps of Gaussian Quadrature approximation (11 quadrature nodes) [44, 45] to obtain the marginal likelihood and then the Hessian matrix for fixed effects to compute the standard errors (SEs) and p-values for inference. The revised MAP algorithm of the two-cluster version is formulated as below and vectors/matrices are in bold.

E-step

Let t denotes the index of iteration. Given $\omega^{(\mathbf{t})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\nu}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_\mathbf{1}}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_\mathbf{2}}, \mu_1^{(t)}, \mu_2^{(t)}, \pi^{(t)}$, we computed the group membership posterior probability vector $\gamma_\mathbf{1}^{(\mathbf{t})}, \gamma_\mathbf{2}^{(\mathbf{t})}$ (Equation (4.4)) which served as weights to compute the weighted estimates in M-step.

$$\gamma_\mathbf{1}^{(\mathbf{t})} = P(C_i = 1|Y_i, X_i, W_i) = \frac{\pi^{(t)}\mathcal{N}(\omega^{(\mathbf{t})}; \mu_\mathbf{1}^{(\mathbf{t})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_\mathbf{1}})}{\pi^{(t)}\mathcal{N}(\omega^{(\mathbf{t})}; \mu_\mathbf{1}^{(\mathbf{t})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_\mathbf{1}}) + (1-\pi^{(t)})\mathcal{N}(\omega^{(\mathbf{t})}; \mu_\mathbf{2}^{(\mathbf{t})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_\mathbf{2}})} \tag{4.4}$$

Note that $\gamma_\mathbf{2}^{(\mathbf{t})} = \mathbf{1} - \gamma_\mathbf{1}^{(\mathbf{t})}$

M-step

(1) Given $\nu^{(\mathbf{t})}, \omega^{(\mathbf{t})}$, we maximized the conditional log likelihood $\sum_i lnf(\epsilon_\mathbf{i} \mid \nu_i^{(t)}, \omega_i^{(t)})$ and obtained the estimates of $\beta^{(\mathbf{t+1})}, \tau^{(\mathbf{t+1})}$ using Quasi-Newton algorithm.

(2) Given $\beta^{(\mathbf{t+1})}, \tau^{(\mathbf{t+1})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\nu}, \omega^{(\mathbf{t})}$, we computed the gradient $\mathbf{G}_{\nu}$ and the Hessian matrix $\mathbf{H}_{\nu}$ of the joint log likelihood function. Then we updated ν by a Newton step and update $\sigma^{\mathbf{2}(\mathbf{t})}_{\nu}$ using the formula below.(Equation 4.5)

$$
\begin{aligned}
\nu^{(\mathbf{t+1})} &= \nu^{(\mathbf{t})} - \mathbf{H}_{\nu}^{(\mathbf{t})^{\mathbf{T}}} \mathbf{G}_{\nu}^{(\mathbf{t})} && (4.5)\\
\sigma^{\mathbf{2}(\mathbf{t})}_{\nu} &= \frac{1}{N}\{\nu^{(\mathbf{t+1})^{\mathbf{T}}} \nu^{(\mathbf{t+1})} + \mathbf{1}^{T} \mathbf{H}_{\nu}^{(\mathbf{t})^{-1}} \mathbf{1}\} && (4.6)
\end{aligned}
$$

(3) Given $\beta^{(\mathbf{t+1})}, \tau^{(\mathbf{t+1})}, \gamma_{\mathbf{1}}^{(\mathbf{t})}, \gamma_{\mathbf{2}}^{(\mathbf{t})}, \sigma^{\mathbf{2}(\mathbf{t})}_{\nu}, \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_{\mathbf{1}}} \sigma^{\mathbf{2}(\mathbf{t})}_{\omega_{\mathbf{2}}}, \mu_{\mathbf{1}}^{(\mathbf{t})}, \mu_{\mathbf{2}}^{(\mathbf{t})}$ we computed $\mathbf{H}_{\omega}$ and $\mathbf{G}_{\omega}$ and updated ω by a Newton step (Equation 4.7),

$$
\omega^{(\mathbf{t+1})} = \omega^{(\mathbf{t})} - \mathbf{H}_{\omega}^{(\mathbf{t})^{-\mathbf{1}}} \mathbf{G}_{\omega}^{(\mathbf{t})}, \tag{4.7}
$$

and then updated ${\sigma^{\mathbf{2}}}_{\omega_{\mathbf{1}}}^{(\mathbf{t})}, {\sigma^{\mathbf{2}}}_{\omega_{\mathbf{2}}}^{(\mathbf{t})}, \mu_{1}^{(t)}, \mu_{2}^{(t)}, \pi^{(t)}$ using the weighted formulas (Equation 4.8) below,

$$
\begin{aligned}
\mu_{\mathbf{1}}^{(\mathbf{t+1})} &= \frac{1}{N_1} \mathbf{1}^{T} \mathcal{DIAG}(\gamma_{\mathbf{1}}^{(\mathbf{t})}) \omega^{(\mathbf{t+1})},\\
\mu_{\mathbf{2}}^{(\mathbf{t+1})} &= \frac{1}{N_2} \mathbf{1}^{T} \mathcal{DIAG}(\gamma_{\mathbf{2}}^{(\mathbf{t})}) \omega^{(\mathbf{t+1})},\\
\sigma^{\mathbf{2}(\mathbf{t+1})}_{\omega_{\mathbf{1}}} &= \frac{1}{N_1}\{(\omega^{(\mathbf{t+1})} - \mu_{\mathbf{1}}^{(\mathbf{t+1})})^{\mathbf{T}} \mathcal{DIAG}(\gamma_{\mathbf{1}}^{(\mathbf{t})})(\omega^{(\mathbf{t+1})} - \mu_{\mathbf{1}}^{(\mathbf{t+1})}) + \mathbf{1}^{\mathbf{T}} (\mathbf{H}_{\omega}^{(\mathbf{t})})^{-\mathbf{1}} \gamma_{\mathbf{1}}^{(\mathbf{t})}\}\\
\sigma^{\mathbf{2}(\mathbf{t+1})}_{\omega_{\mathbf{2}}} &= \frac{1}{N_2}\{(\omega^{(\mathbf{t+1})} - \mu_{\mathbf{2}}^{(\mathbf{t+1})})^{\mathbf{T}} \mathcal{DIAG}(\gamma_{\mathbf{2}}^{(\mathbf{t})})(\omega^{(\mathbf{t+1})} - \mu_{\mathbf{2}}^{(\mathbf{t+1})}) + \mathbf{1}^{\mathbf{T}} (\mathbf{H}_{\omega}^{(\mathbf{t})})^{-\mathbf{1}} \gamma_{\mathbf{2}}^{(\mathbf{t})}\}
\end{aligned}
\tag{4.8}
$$

And,

$$\pi^{(t+1)} = \frac{N_1}{(N_1 + N_2)} \tag{4.9}$$

$\mathcal{DIAG}$ is a function to convert a vector to a diagonal matrix so that the matrix diagonals are the vector entries. $N_1 = sum(\gamma_1^{(t)})$ and $N_2 = sum(\gamma_2^{(t)})$ and $N = N_1 + N_2$. Then we repeated E-step and M-steps until convergence criterion was met.

For estimation, the EM-MAP algorithm can provide unbiased solutions for all parameters. But for inference, we may still have to estimate the standard errors so that the maximum likelihood method was used following the EM-MAP step. The EM-MAP step produced feasible initial values and also the unbiased estimate of the marginal membership probability. With the known marginal membership probability, the proposed model became identifiable. Using the parameter estimates from the EM-MAP step as the initial values which were sufficiently close to the solutions, the maximum likelihood step took much fewer iterations to converge. Given the membership probability, Gauss-Hermite quadrature with 11 quadrature points was used to approximate the marginal likelihood and then yield the Hessian inverse of $\beta, \tau, \mu_{\mathbf{1}}, \mu_{\mathbf{2}}$ when the optimization objective achieved the optima. The corresponding standard errors were calculated as the square root of the diagonal entries in the Hessian inverse. To simplify, in the following analyses, we assumed random subject effects were independent across subjects and therefore $\sigma^{\mathbf{2}}{}_{\nu}, \sigma^{\mathbf{2}}{}_{\omega_{\mathbf{1}}}, \sigma^{\mathbf{2}}{}_{\omega_{\mathbf{2}}}$ are diagonal matrices with $\sigma^2_{\nu}, \sigma^2_{\omega_1}, \sigma^2_{\omega_2}$ as diagonal elements respectively.

As an initial value finding strategy, the convergence criterion for EM-MAP step was that the relative change of the conditional log likelihood given the random effects was set to $< 1e-6$. If the convergence criterion was not met after 50 iterations, then the optimization would discontinue.

4.4 Simulation

We tested the proposed model in scenarios parametrized by three combinations of π, μ_1, μ_2: (a) $\pi = 0.9, \mu_1 = 0, \mu_2 = -10$; (b)$\pi = 0.6, \mu_1 = 0, \mu_2 = -10$; (c) $\pi = 0.9, \mu_1 = 0, \mu_2 = -4$. Scenario (a) mimics the real scenario that only small portion of the data are longitudinally identical. Two clusters are separated faraway apart and one of the cluster is of small size. The sizes of clusters in Scenario (b) are more balanced than in Scenario (a). Comparing to Scenario (a)(b), in Scenario (c), two clusters are closer to each other and the true means of both clusters are not extremely negative. This setting mimics a more general scenario where the within-subject assessments are still variable but display different levels of variability. Within each simulated dataset, there are 200 subjects and each subject has 10 within-subject occasions.

Thus, the data will be generated as follows:

(1) Generate a latent binary variable U_i for each subject from a Bernoulli distribution with probability π.

(2) If $U_i = 1$ then we sample ω_i from $\mathcal{N}(\mu_1, \sigma^2_{\omega_1})$ but if $U_i = 0$ then sample ω_i from $\mathcal{N}(\mu_2, \sigma^2_{\omega_2})$, given $\sigma^2_{\omega_1}$ and $\sigma^2_{\omega_2}$.

(3) Then sample ν_i from $\mathcal{N}(0, \sigma^2_{\nu})$

(4) Generate $\mathbf{X_{ij}}$ as the observed covariates.

(5) Given $\mathbf{X_{ij}}$, ν_i, ω_i, simulate Y_{ij} from a normal distribution parametrized by a two-level Mixed Effect Location-Scale model (Equation (4.1)).

The true latent binary variable U was excluded from the dataset and presumed to be unknown. The performance of the conventional MELS and proposed model were compared in

Table 4.1: Simulation Results

	Scenario (a)			Scenario (b)			Scenario (c)				
		MELS-MS			MELS-MS			MELS		MELS-MS	
Param	**True**	Mean	CR	**True**	Mean	CR	**True**	Mean	CR	Mean	CR
$\boldsymbol{\beta_0}$	1	1.00	.97	1	1.00	0.94	1	.53	.06	1.00	.96
$\boldsymbol{\beta_1}$	2	2.00	.94	2	1.99	0.94	2	2.00	.82	2.01	.95
$\boldsymbol{\tau_1}$	-2	-2.00	.93	-2	-2.00	0.94	-2	-1.84	.23	-2.00	.93
$\boldsymbol{\mu_1}$	0	.00	.93	0	-.01	0.96	0	-1.41	.08	.01	.89
$\boldsymbol{\mu_2}$	-10	-9.99	.87	-10	-9.99	.93	-4			-3.71	.68
$\boldsymbol{log(\sigma^2_\nu)}$	0	-.02	.96	0	-.00	.96	0	.10	.33	-.01	.94
$\boldsymbol{log(\sigma^2_{\omega_1})}$	0	-0.02	0.94	0	-0.05	0.93	0	3.11	.00	-.09	.88
$\boldsymbol{log(\sigma^2_{\omega_2})}$	0	-0.14	0.89	0	-0.04	0.95	0			-.08	.78
$\boldsymbol{\pi}$	0.9	0.90		0.6	0.60		0.9			.87	

Param: Parameters;
True: True value;
CR: coverage rate;
MELS: conventional mixed effect location-scale model;
MELS-MS: proposed mixed effect location-scale model with mixture distributed scale random effects.

terms of bias from the true values, the coverage rate of the parameters of interest, and the number of successfully converged solutions of the EM-MAP step (Table 4.1).

We compared the performance of the proposed model, mixed effect location-scale model with mixture distributed scale effects (MELS-LS) under three different simulated scenarios. The proposed model recovered all the parameters with small bias and provided correct coverage of about 0.95 for all the fixed effect parameters. The proposed model performed the best in scenario (b) as the number of convergent solutions ($N_{convg} = 178$ out of 200) is the highest among the three scenarios and all of the model parameters were estimated with correct coverage and and reasonable bias. For scenario (a) ($N_{convg} = 152$ out of 200), the size of the second cluster is relatively small, which may not provide sufficient sample size to estimate the cluster-specific parameters. We can see that μ_2 and $\sigma^2_{\omega_2}$ were estimated with slightly lower coverage rate than 90%. Thus, balanced sample size and more distinguishable clusters resulted in more faster convergence.

In scenario (c) where two clusters are close to each other and become less distinguishable, the coverage rates for both μ_1, μ_2 and $\sigma^2_{\omega_1}$, $\sigma^2_{\omega_2}$ were all lower than 90% and substantial biases for μ_1 and μ_2 emerged. In addition, less distinguishable clusters may also lead to bias in estimating π_k as the EM-MAP step can hardly converge with the pre-defined convergent criterion. Again, as an initial value finding strategy, the EM-MAP algorithm didn't have to converge. But without convergence, the estimated and potentially biased π_k may further bias the cluster mean and variability estimates.

Thus, the proposed model applied in scenario (a)(b) provided better performance in terms of estimation and inference than in scenario (c).

We also applied the conventional MELS to these three scenarios. In Scenario (a)(b) with subjects had low variance, the conventional MELS can hardly coverage for the EM-MAP step and the estimated $\boldsymbol{\omega}$ went extreme. Thus we conclude that conventional MELS is very sensitive to even small number of subjects with low variability. The conventional MELS can only converge on data simulated from Scenario (c). The number of successfully convergent solutions for conventional MELS applied on Scenario (c) was 79 out of 200. But In terms of biases and coverage rates, β_0 and β_1 didn't have much bias and the coverage rates were around 95%. But severe biases and low coverage for τ_1 emerged as a result of mis-specification for random scale effect distribution.

4.5 Application on MATCH Dataset

A motivating example with identical within-subject responses is the MATCH (Mother And Their Children's Health) Study [17] which used EMA technology to explore the within-day association between maternal stress and the subject's dietary intake and physical activity level. In this analysis, the extent of "feeling happy" was the primary outcome in the analysis, which was scored from 1-5 by increment of 1. One challenge to analyze these data is that 8 out of 144 subjects provided identical entries of happiness at wave 5. There are 2 to 34 within-subjects measurements across the study. Specifically for number of measurements within those 8 subjects, the maximum was 33 and minimum number was 2. The identical entries were of values 2, 4 and 5. We hypothesized that the mean and variability of subject's happiness are associated with the level of vigorous physical activity cumulatively measured in the 120 minutes interval before the measurement (ActBef). Activity level was measured by actigraphy counts, which were highly right-skewed so we first transformed this variable to log scale using the function $log(x + 1)$.

4.5.1 Application 1: Variance Pattern Clustering

The first usage of the proposed model is to classify subjects' variance patterns. We first excluded those 8 subjects with identical responses. For the rest 136 subjects with varying responses for happiness, we hypothesized that there were two latent clusters to describe subjects' variance patterns. We first applied the conventional MELS assuming single cluster and then the MELS-MS with two clusters on the reduced sample ($N = 136$) and compared the two candidate models in terms of model estimates and -2 log likelihood (-2LL) and Bayesian Information Criterion (BIC) (Table 4.2).

For the reduced sample ($N = 2931$), β_{ActBef} is significant and positive as the 95% confidence interval didn't include zero for both candidate models. It means vigorous activity before the assessment is significantly related to higher level of happiness. Although τ_{ActBef} is not significant, vigorous activity is still positively associated with more variability in happiness.

Again, conventional MELS is a special case of our proposed model with only one single cluster. From Table 4.2, the mean of random scale effects in MELS is between μ_1 and μ_2 in MELS-MS. As the mean of random scale effects in conventional MELS can be viewed as the weighted average of random scale effects from the two clusters. For $\sigma^2_{\omega_1}$ and $\sigma^2_{\omega_2}$ in the proposed model, cluster 1 is much more concentrated than cluster 2, which means cluster 1 subjects are more similar to one another than cluster 2 in terms of within-subject variability, after controlling for the activity effect. For subjects from cluster 1, the range of standard deviation of the outcome was (.659, 2.049) and for cluster 2, the range was (.229, .630). Corresponding to estimates of μ_1 and μ_2, we can see that cluster 1 subjects had dominantly higher variability than cluster 2 subjects. Based on the BIC value, MELS-MS had better performance than MELS for this reduced sample.

Table 4.2: Variance Pattern Clustering ($N_{sub} = 136$)

	MELS		**MELS-MS**	
Parameters	**Est.**	**95% CI**	**Est.**	**95% CI**
$\boldsymbol{\beta_{Intercept}}$	3.189	(3.054, 3.323)	3.188	(3.051,3.325)
$\boldsymbol{\beta_{ActBef}}$	.025	(.001, .048)	.030	(.006, .053)
$\boldsymbol{\tau_{ActBef}}$	.011	(-.027,.049)	.023	(-.018,.063)
$\boldsymbol{\mu_1}$	-.089	(-.207, .028)	.070	(-.005,.145)
$\boldsymbol{\mu_2}$			-1.244	(-1.506, -.982)
$\boldsymbol{\sigma^2_{\nu}}$	.585	(.454, .756)	.614	(.476,.792)
$\boldsymbol{\sigma^2_{\omega_1}}$	.340	(.243, .476)	.023	(.004, .120)
$\boldsymbol{\sigma^2_{\omega_2}}$			.502	(.214, 1.180)
$\boldsymbol{\pi}$			.838	
$\boldsymbol{-2LL}$	8584.7		8493.7	
$\boldsymbol{BIC}$	8632.6		8565.5	

Est: Estimate
CI: Confidence Interval
MELS: conventional mixed effect location-scale model;
MELS-MS: proposed mixed effect location-scale model with mixture distributed scale random effects.
-2LL: -2 log likelihood
BIC: Bayesian Information Criterion

4.5.2 Application 2: modelling data with identical within-subject responses

In the second application, perfectly identical within-subject responses were included. For the reduced sample (Table 4.2), there may exist two clusters in variability as the two-cluster MELS-LS had the best performance. Thus, for the full sample ($N = 3021$), we assumed there may exist three clusters in subjects' variance patterns. We assumed subjects in cluster 1 provided varying responses of happiness and those in cluster 2 provided responses of low variance and subjects from cluster 3 provided perfectly identical responses.

When using the whole sample, conventional MELS can barely converge. One solution for handling identical within-subject outcomes is to add negligible random variance to the outcomes. But even with the added variance, the estimation of MELS was still unstable but our proposed model successfully converged and identified the cluster memberships with high accuracy.

Corresponding to the introduction section, we compare different strategies to deal with the identical within-subject responses: (i) to exclude subjects with identical responses (Table 4.2); (ii) to retain those responses and add small amount of random variance (exp(-10)) so that the within-subject variance can still be scaled to sufficiently close to zero (Table 4.3).

Those 8 subjects with identical responses were of high leverage and their scale density functions can be very spiky even after some turbulence had been added to these responses. We want to note that these spiky log densities can be greater than 0 so that after including these 8 subjects, the *-2LL* value can even become smaller. For the fixed effects, the activity effect on the mean happiness level, β_{ActBef}, became insignificant after the identical responses were included. Although being insignificant, fixed effects τ_{ActBef} in all three models were consistently positive, which means the level of vigorous activities before assessment was associated

Table 4.3: Modelling Identical Within-Subject Responses ($N_{sub} = 144$)

	MELS		**MELS-MS**		**MELS-MS**	
Parameters	**Est.**	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**
$\beta_{Intercept}$	3.325	(3.082, 3.349)	3.218	(3.078, 3.358)	3.219	(3.079,3.359)
β_{ActBef}	-.001	(-.003, .002)	.002	(.000,.003)	-.001	(-.003, .001)
τ_{ActBef}	.037	(.002, .073)	.026	(-.011,.063)	.027	(-.010,.064)
μ_1	.181	(.120, .242)	-.147	(-.248, -.045)	.079	(.004,.155)
μ_2			-6.690	(-6.943,-6.438)	-1.184	(-1.468, -.900)
μ_3					-9.937	(-10.222, -9.652)
σ^2_ν	.608	(.474, .781)	.691	(.537,.956)	.692	(.538, .890)
$\sigma^2_{\omega_1}$	3.456	(3.198, 3.736)	.244	(.172, .347)	.018	(.002, .129)
$\sigma^2_{\omega_2}$			15.455	(13.457,17.750)	.128	(.019, .879)
$\sigma^2_{\omega_3}$					13.530	(11.665, 15.694)
π_1			.890		.764	
π_2			.110		.173	
π_3					.063	
$-2LL$	8316.6		7934.8		7921.0	
BIC	8364.6		8006.9		8017.2	

True: True value;
CR: coverage rate;
SD: standard deviation;
MELS: conventional mixed effect location-scale model;
MELS-MS: proposed mixed effect location-scale model with mixture distributed scale random effects.
-2LL: -2 log likelihood
BIC: Bayesian Information Criterion

with greater within-subject variability.

Based on the -2 log likelihood value, the three-cluster MELS-MS performed the best on the full sample. When using the three-cluster version, including the subjects with identical responses didn't affect the cluster 1 and cluster 2 parameters much. As was observed, μ_1 and μ_2, $\sigma^2_{\omega_2}$ and $\sigma^2_{\omega_1}$ estimated on the whole sample were close to those estimated on the reduced sample. And the estimated marginal proportion of subjects in cluster 3 was .063, which was close to 8/144=.056, the observed proportion of subjects with identical responses. But there was one subject (ID 11065) misclassified into cluster 3. Among her 17 happiness assessments, 16 out of 17 were rated as 4 and only one was rated as 5, which were near identical.

However, if we compared the MELS-MS on the whole sample (Table 4.3) and conventional

MELS with single cluster fitted on the reduced sample (Table 4.2, cluster 1 parameter estimates were substantially different. Cluster 2 for the MELS-MS on the whole sample didn't exclusively include the subjects with identical responses but 7 other subjects with low within-subject variance were mixed into cluster 2. For those 7 subjects, the standard deviation of the outcome was ranged as (.229, .511). The more clusters we specified, the less influence of identical responses on the original existing clusters. But according to BIC value, the two-cluster MELS-MS achieved better balance between goodness of fit and model complexity then the three-cluster version.

4.6 Discussion

In this article, we addressed a prevalent issue of the (near) identical responses in intensive longitudinal surveys at the data analysis stage post the survey design. As the subjects with identical responses had high leverage, weighting likelihood approach may be needed in the future. In addition, to added negligible noises to the identical responses can be analogous to random imputation strategy. Future research can focus on the adjustment for the randomness brought from these added noises.

Our proposed method can be used even when no information regarding the data validity is provided and it is able to include all available data for modelling. Given such data, we framed this issue as a clustering problem for variability and assumed all available data were valid real data. In the generalized linear mixed model class, the mixed effect location-scale modelling framework allows flexible scale random effect distributions in the within-subject variance function. Actually the Gaussian mixture distribution for scale random effects is not perfectly applicable for identical within-subject responses but we still used Gaussian mixture distribution for its convenience to construct the continuous likelihood density function for estimation and for better interpretability. Other than Gaussian mixture, as we have talked in the introduction section, there can still be other non-parametric/semi-parametric specifications for random effect distributions which can be more applicable for modeling perfectly identical responses.

Other than to assume the data are valid, there is another possibility of the so-called 'careless responses' [38] in survey data, which is an issue in the similar sense to missing data. Once we have information to verify the invalidity of these identical responses, we can treat these data as missing data. Thus, it can be argued that concepts as well as methods from missing data can be transplanted to the scenario of 'careless responses'[38]. For the real data analysis example, we can possibly use a technique similar to multiple imputation by imputing small variances to the identical responses multiple times. To deal with the introduced randomness, we can implement this approach multiple times and aggregated the estimates using Rubin's rule which is often used in combining estimates after multiple imputations for missing data[48]. Other than imputation, some weighting methods for handling non-response bias may also be necessary for 'careless responses'. We can arbitrarily set the weights for these identical responses as 1/100 of the rest of the responses. But other more flexible and informative strategies for setting the weights can also be used [52, 36]. In addition, analogous to informative missingness, the probability of having identical responses may not be independent of the outcome level. For example, we often found that identical responses are of the either highest or lowest value.

The proposed model present in this article was a two-level (between subject-level and within-subject level) and single-outcome version but it can also be extended to multiple-outcome version and three-level version (e.g., subject-level, day-level and assessment-level). For a three-level version, variance clustering can be on the day-level and we can even model the transition between clusters for each day so that booster strategies can be designed and applied to stimulate subjects' compliance to carefully respond to the EMA items. And for the multiple-outcome version, composite information regarding the within-subject variability in other outcomes can help to decide whether the identical entries are "careless responses". In conclusion, the proposed methodology is practical and has great potential for data classification and data screening for intensive longitudinal survey data.

CHAPTER 5

A NEGATIVE BINOMIAL MIXED EFFECTS LOCATION-SCALE MODEL FOR PHYSICAL ACTIVITY DATA PROVIDED BY WEARABLE DEVICES

5.1 Introduction

Accompanied with recent technological advances, the use of wearable devices, e.g., accelerometers and smart watches, have become increasingly prevalent [61, 8]. Studies have been shown that wearable devices not only provide more accurate tracking for subjects' physical activity (PA) levels such as step counts or minutes in moderate to vigorous intensity physical activities (MVPA) but also better engage subjects into physical activities and further to improve the general health of the population [7, 11].

PA data collected by these wearable devices are often real-time stamped and highly intensive. For example, in the example presented later [17, 35], the accelerometer data were captured in 30-sec or 1-min epochs and then each epoch was classified and labelled as "sedentary","light-intensity", "moderate-intensity", or "rigorous-intensity." These were then summarized for each hour, within a day and across days. Also, PA data are often non-negative and integer-valued and can include an excessive number of zero counts, e.g., for most of the day, subjects might have only light-intensity PA or even sedentary without any MVPA. This can result in severely skewed outcome distributions [4], and so statistical models based on a Gaussian assumption may not be applicable. Popular parametric methodologies for modeling such longitudinal count data include the Poisson mixed model which assumes no over-dispersion and that the mean is equal to the variance [4], or the Negative Binomial mixed model when over-dispersion is present, i.e., variance is larger than the mean, [64, 67]. Corresponding zero-inflated versions have been developed to account for the excessive zeros across time

[63]. Most of these publications have focused on modeling the mean level of counts. However, for these intensive longitudinal PA data collected by wearable devices, different subjects may display different mean levels as well as dispersion levels. In this case, the conventional generalized linear mixed model where the dispersion parameter is assumed to be a constant parameter across all subjects may not be applicable.

The mixed effect location-scale (MELS) model [23, 24] is an intuitive modeling framework which enables subject-level random effects in both the mean and variance. MELS models for continuous outcomes [23] as well as ordinal outcomes [24] have been developed but thus far have not been developed for modeling the subject-specific mean level and dispersion level in intensive longitudinal counts. A possible choice for modeling longitudinal count data is to extend the Poisson mixed model into a location-scale modeling framework. But because of its assumption of equal mean and variance, the mean and variance of the Poisson model cannot be decoupled. A second choice is the negative binomial mixed effects model where a separate parameter is specified for handling the over-dispersion. Thus, we propose a MELS model assuming the count outcome follows a negative binomial distribution (NBLS). In our proposed model, we specify a location random effect in the mean function that distinguishes different subject-level PA averages, and a scale random effect in the dispersion function that reflects different levels of over-dispersion and then the variability across subjects. These random effects of location and scale characterize unobserved subject-specific physical activity patterns. For example, a subject with a regular activity schedule and more consistent PA amount over time might be a less dispersed subject but on the other hand, a subject who exercises excessively on one single day would be a more dispersed subject.

To account for excessive zeros in the counts, we can extend the proposed NBLS as a hurdle NBLS model (H-NBLS) or a zero-inflated NBLS model (ZI-NBLS). The only difference between the hurdle approach and zero-inflated approach is the distributional assumption for the zero counts. We will elaborate on this difference in the method section. For both, the non-zero counts are assumed to follow the same zero-truncated negative binomial distribution. The hurdle approach is sometimes preferred over the zero-inflated approach for better computational efficiency [40]. In this article, we will first compare the proposed location-scale models to the conventional negative binomial (hurdle) mixed model via a simulation study. Then we will apply the proposed model to a real dataset, and show how the subject-level scale random effects improve the fit to the real dataset.

5.2 The Proposed Method

For modeling the dispersed counts, we will use a negative binomial distribution. The negative binomial distribution is often expressed as $\mathcal{NB}(r, p)$ (Equation 5.1), with r as the over-dispersion parameter which is positive, and $p = \frac{r}{r+\mu}$ where μ is the mean of the distribution. While the variance σ^2 of the Poisson distribution is equal to μ, in the negative binomial distribution $\sigma^2 = \mu + \frac{1}{r}\mu^2$. The larger r is, the less over-dispersion is present. In other words, as r gets large, the negative binomial distribution approximates the Poisson distribution. In a conventional negative binomial mixed effect model, the over-dispersion parameter is a fixed parameter across all subjects. But given the nature of the intensive longitudinal PA data, we propose to extend this and allow varying dispersion levels for each subject.

$$\mathcal{NB}(y; r, p) = \binom{y + r - 1}{y} (1 - p)^r p^y \tag{5.1}$$

5.2.1 Model 1: Negative Binomial Mixed Effect Location-Scale Model (NBLS)

Let i indexes subjects ($i = 1, \ldots, N$) and j the observations within subjects ($j = 1, \ldots, n_i$), then the model can be written as:

$$\begin{aligned} \mu_{ij} &= \exp(X_{ij}\beta + \nu_i), & (5.2) \\ r_{ij} &= \frac{1}{\exp(W_{ij}\tau + \omega_i)}, & (5.3) \end{aligned}$$

with the location and scale random effects distributed as:

$$\begin{pmatrix} \nu_i \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma_\omega^2 \end{pmatrix}\right] \tag{5.4}$$

Note that ν_i and ω_i can be correlated, i.e., $\sigma_{\nu\omega}$ can be non-zero valued. Thus, the variance of the outcome is expressed as in Equation 5.5.

$$\sigma_{ij}^2 = \mu_{ij} + \exp(W_{ij}\tau + \omega_i)\mu_{ij}^2 \tag{5.5}$$

We assume the count outcome $Y_{ij} \sim \mathcal{NB}(r_{ij}, p_{ij})$ where $p_{ij} = \frac{r_{ij}}{r_{ij}+\mu_{ij}}$. And X_{ij} and W_{ij} are the covariate vectors at observation j within subject i in the mean and dispersion component respectively. When all ω_i are consistently zero, the proposed NBLS model degenerates to a negative binomial mixed model (NB-Mixed).

With the above specifications, we can write the marginal likelihood function, which will be utilized for the maximum likelihood estimation approach. (Equation 5.6)

$$\mathcal{L}(\beta, \tau, \alpha, \sigma_\nu^2, \sigma_\omega^2) = \int_\nu \int_\omega \{\prod_{i=1}^{N} \{\prod_{j=1}^{n_i} \mathcal{NB}(y_{ij}; p_{ij}, r_{ij})\} \\ \mathcal{N}(\nu_i; 0, \sigma_\nu^2)\mathcal{N}(\omega_i; 0, \sigma_\omega^2) d\nu_i d\omega_i\}. \tag{5.6}$$

5.2.2 Model 2: Zero-Inflated Negative Binomial Mixed Effect Location-Scale Model (ZI-NBLS)

Zero-inflated and hurdle models can be used to deal with the issue of excessive zero counts, but the ways are different. Model 2 (zero-inflated) assumes the source of zeros come from two sources. For example, if the subject didn't wear the device validly, the minutes of moderate exercises would not be recorded and the outcome would definitely be zeros; but if the subject wore the device but didn't have any PA, then the outcome would also be zero but this zero comes from the negative binomial distribution. Therefore, Model 2 (Equation 5.7, 5.8) specifies a mixture of point mass at zero and a negative binomial distribution to model the outcomes.

Assume that the probability of being in the negative binomial component is π_{ij}, where π_{ij} can further be expressed as $\pi_{ij} = \frac{1}{1+\exp(-(A_{ij}\alpha+\psi_i))}$, which is the logistic response function. Here, the vector A_{ij} represents the observed covariates and ψ_i is the subject-level random intercept in the logistic component. Thus, the higher ψ_i is, the higher the probability of being in the negative binomial component.

$$P(Y_{ij} = 0|\nu_i, \omega_i) = (1 - \pi_{ij}) + \pi_{ij}\mathcal{NB}(0; r_{ij}, p_{ij}) \tag{5.7}$$

$$P(Y_{ij} = y|Y_{ij} > 0, \nu_i, \omega_i) = \pi_{ij}\mathcal{NB}(y; r_{ij}, p_{ij}) \tag{5.8}$$

This provides the marginal likelihood function that will be used for the maximum likelihood estimation approach, detailed later (Equation 5.11).

5.2.3 Model 3: Negative Binomial Location-Scale Hurdle Mixed Model (H-NBLS)

While the zero-inflated model assumes the source of zeroes to be a mixture, the hurdle model combines the two sources and adopts a two-stage modeling approach. First, it posits a logistic mixed model to model the probability of the outcome to be non-zero. Then it uses a truncated negative binomial to model the non-zero counts.

Similar to Model 2, but now, $\pi_{ij} = \frac{1}{1+\exp(-(A_{ij}\alpha+\psi_i))}$ is the probability of being non-zero rather than being in the negative binomial component, and naturally $1-\pi_{ij}$ is the probability of being zero, regardless of the source of the zeros.

$$P(Y_{ij} = 0) = 1 - \pi_{ij} \tag{5.9}$$

$$P(Y_{ij} = y | Y_{ij} > 0) = \pi_{ij} \frac{\mathcal{NB}(y; r_{ij}, p_{ij})}{1 - \mathcal{NB}(0; r_{ij}, p_{ij})} \tag{5.10}$$

The form of the marginal likelihood function for H-NBLS is the same as Equation 5.11.

$$\mathcal{L}(\beta, \tau, \alpha, \sigma_\nu^2, \sigma_\omega^2, \sigma_\psi^2) = \int_\psi \int_\omega \int_\nu \prod_{i=1}^{N} \{ (\prod_{j:y_{ij}=0}^{n_i} P(Y_{ij} = 0) \prod_{j:y_{ij}\neq 0}^{n_i} P(Y_{ij} = y | Y_{ij} > 0)) \mathcal{N}(\nu_i; 0, \sigma_\nu^2) \} \mathcal{N}(\omega_i; 0, \sigma_\omega^2) \mathcal{N}(\psi_i; 0, \sigma_\psi^2) d\nu_i d\omega_i d\psi_i \}. \tag{5.11}$$

We can see that when the probability that $\mathcal{NB}(0; r_{ij}, p_{ij})$ is small, ZI-NBLS will approach

H-NBLS. However, the over-dispersion parameter has different meaning between ZI-NBLS and H-NBLS. As in Model 2, part of the zeros contribute to the estimation of r and μ, but in Model 3 the over-dispersion parameter is estimated using only the strictly positive values.

5.3 Estimation and Inference

All of the models specified can be estimated using maximum likelihood method, where the joint likelihood function will be first integrated over the random effects ν, ω, ψ, yielding the marginal likelihood function (Equation 5.6,5.11) as the optimization objective. Typically, the integral cannot be expressed in closed form. However, since the random effects are assumed to follow the Gaussian distribution, the adaptive Gauss-Hermite quadrature method can be used for the integral approximation. As this step can be time consuming, one can perform this integration using one quadrature point which is analagous to the Laplace Approximation. For this, SAS PROC NLMIXED can be used, including specification of the Newton-Raphson algorithm for the optimization.

5.4 Simulation Studies

We first simulated 100 datasets respectively using (1) NBLS, (2) ZI-NBLS, and (3) H-NBLS and validated these three proposed models using the simulated datasets (Table 5.1). The goal for this simulation study is to validate the proposed models and compare the computational efficiency between H-NBLS and ZI-NBLS. Given this set of parameter values (Table 5.1, 5.2), the mean and variance of the simulated outcomes were 37.45 and 63872.67 across all the datasets and if we only included the > 0 outcomes, the mean and variance were 50.54 and 85547.68, respectively. The only covariate corresponding to $\beta_1, \tau_1, \alpha_1$ in this simulation study was simulated using a standard normal distribution which result in 26% zero-valued outcomes.

From Table 5.1, all of the three proposed models yielded coverage rates of around 95% and obtained parameter estimates with reasonable low levels of bias. However, the ZI-NBLS is more computationally intensive than H-NBLS as ZI-NBLS modeled the zero counts using a mixture distribution and the ZI-NBLS approach was more sensitive for the initial values and took longer to converge during the simulation study. As was introduced in Section 5.2, the non-zero part of ZI-NBLS and H-NBLS are the same and both are assumed to follow a zero-truncated negative binomial distribution and therefore the β and τ estimates should be very close. Given the greater efficiency of H-NBLS, for the following analyses, we mainly modeled the zero counts using the hurdle approach.

In the second simulation study, we simulated the data from the H-NBLS model, and we compared the conventional approaches (NB mixed model and the Hurdle negative binomial mixed model(H-NB-Mixed)), to the proposed approaches (the NBLS model and the H-NBLS model). We explored how the added scale random effect and also the hurdle component affect the estimation biases as well as the coverage rates. The coverage rate was calculated as the proportion of times that the 95% confidence intervals covered the true values (Table 2).

From data simulated from H-NBLS, all of the less complex candidate models (NB-Mixed, H-NB-Mixed and NBLS) experienced issues in terms of estimation and inference. This is noteworthy since the NB-mixed and H-NB-mixed represent popular choices in modeling multilevel count data with over-dispersion. Of these two, the NB-Mixed model had the worse performance in terms of coverage rates and biases. Comparing H-NB-mixed to H-NBLS, if the scale random effect ω_i was suppressed (i.e., the extra effects other than the observed covariate effects on dispersion were assumed to be consistent across subjects), then both the overall dispersion τ_0 and the subject-level location variance σ_ν^2 were overly estimated.

Table 5.1: Model validation on 100 datasets simulated respectively from NBLS, ZI-NBLS,H-NBLS

	NBLS		**ZI-NBLS**		**H-NBLS**		**True**
Param	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**Value**
$\boldsymbol{\beta_0}$	0.95	2.00(0.10)	0.95	2.01(0.10)	0.96	1.97(0.15)	2
$\boldsymbol{\beta_1}$	0.95	-1.50 (0.02)	0.95	-1.50(0.02)	0.97	-1.51(0.10)	-1.5
$\boldsymbol{\tau_0}$	0.97	-0.01(0.11)	0.93	0.00(0.12)	0.96	0.03(0.26)	0
$\boldsymbol{\tau_1}$	0.94	0.50 (0.04)	0.93	0.50 (0.04)	0.95	0.50(0.14)	0.5
$\boldsymbol{\alpha_0}$			0.94	2.01(0.20)	0.91	1.98(0.17)	2
$\boldsymbol{\alpha_1}$			0.94	-0.48(0.10)	0.96	-0.49(0.04)	-0.5
$ln(\sigma_\nu^2)$	0.99	0.01 (0.14)	0.93	-0.05(0.16)	0.92	0.00(0.27)	0
$ln(\sigma_\omega^2)$	0.95	-0.05(0.18)	0.94	-0.06(0.17)	0.98	-0.07(0.46)	0
$cov(\nu, \omega)$	0.94	0.50(0.12)	0.91	0.47 (0.13)	0.97	0.47(0.21)	0.5
$ln(\sigma_\psi^2)$			0.90	-0.19 (0.26)	0.95	-0.06(0.18)	0

Param: Parameters
CR: coverage rate
SD: Standard Deviation

However, the biases as well as coverage rates for the covariate effects (β_1,τ_1) in the H-NB-Mixed model were still acceptable. Ignoring the excessive zeros (NBLS) can be more detrimental than ignoring the random scale effect (H-NB-mixed) as it resulted in extremely low coverage rates on the parameters and biased the covariate effects even in the opposite direction.

5.5 Application to MATCH data

5.5.1 MATCH Data

The MATCH (Mother And Their Children's Health) Study [17] used Ecological Momentary Assessment (EMA) technology to explore the within-day association between maternal stress and the subject's dietary intake and physical activity level. The study design involved six semi-annual measurement waves across three years. It used an EMA schedule with random prompts approximately every two hours during waking, non-school time weekends and weekdays and subjects were followed for seven days for each wave. Participants in the MATCH study included both the mothers and the children. Here, for our example, we only used moth-

Table 5.2: Model comparison on datasets simulated from H-NBLS: comparing NB-mixed, NBLS, H-NB-mixed, H-NBLS

	NB-Mixed		**NBLS**		**H-NB-Mixed**		**H-NBLS**		**True**
Param	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**CR**	**Mean(SD)**	**Value**
$\boldsymbol{\beta_0}$	0.39	1.79 (0.13)	0.06	1.64 (0.11)	0.96	1.89(0.15)	0.96	1.97(0.15)	2
$\boldsymbol{\beta_1}$	0.00	-0.70(0.03)	0.00	-0.64 (0.03)	0.86	-1.64(0.14)	0.97	-1.51 (0.10)	-1.5
$\boldsymbol{\tau_0}$	0.01	0.58 (0.12)	0.82	-0.11(0.20)	0.25	0.51 (0.21)	0.96	0.03 (0.26)	0
$\boldsymbol{\tau_1}$	0.00	-0.50 (0.10)	0.00	-0.61 (0.09)	0.89	0.59 (0.17)	0.95	0.50 (0.14)	0.5
$\boldsymbol{\alpha_0}$					0.93	2.00 (0.16)	0.91	1.98 (0.17)	-2
$\boldsymbol{\alpha_1}$					0.96	-0.50 (0.05)	0.96	-0.49 (0.04)	0.5
$ln(\sigma_{\nu}^2)$	0.01	-0.77 (0.22)	0.04	-0.85 (0.22)	0.04	0.60 (0.18)	0.92	0.00(0.27)	0
$ln(\sigma_{\omega}^2)$			0.18	0.70 (0.21)			0.98	-0.07 (0.46)	0
$cov(\nu, \omega)$			0.73	0.64 (0.14)			0.97	0.47(0.21)	0.5
$ln(\sigma_{\psi}^2)$					0.95	-0.05(0.18)	0.95	-0.06 (0.18)	0

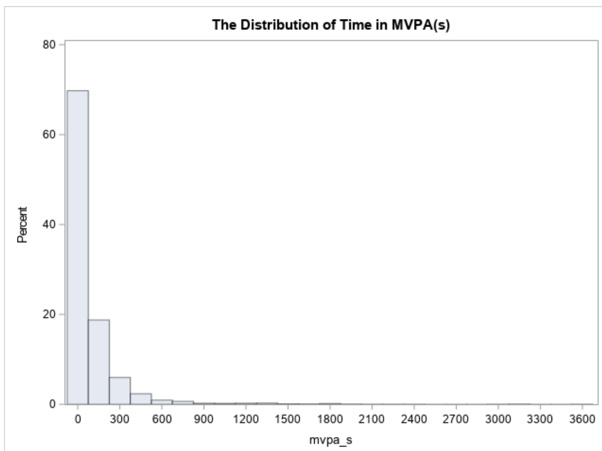
Param: Parameters
CR: coverage rate
SD: Standard Deviation

ers' data from the 4^{th} measurement wave which were considered to be the most compliant subset of subjects. For the 4^{th} wave, the proportion of subjects attending ≥ 5 measurement waves was the highest (70.80%) across all six waves.

PA data in MATCH Study were measured by the Actigraph, Inc. GT3X model accelerometers [17, 35]. In this analysis, we used the seconds in moderate-to-vigorous-intensity activities (MVPA) summarized as 1800 seconds (30 minutes) before and after each randomly prompted EMA assessment as the primary outcome. A reason that we selected this 60-minute window (30 minutes before and after each EMA assessment, altogether 3600 seconds) was to avoid duplicate counts coming from overlapping observation windows as the length of the interval between two consecutive EMA assessments was at least 93 minutes. This outcome was originally recorded in units of minute and non-integer valued. In order to create the integer-valued count-type outcome for modeling, we first transformed the outcome into units of seconds to yield more precise rounding. On this scale, the outcome was highly over-dispersed, as variance (53813.96) was much larger than the mean (97.50 seconds). The distribution was severely right-skewed (Figure 5.1) as the median of this sample was 30 seconds and the 25%

and 75% quantiles were 0 and 90 seconds, respectively. Among the collected 4171 records, 865 of those from 8 subjects were not included in the analysis because of missingness in the outcome as well as in the covariates. The missingness of the outcome variable was due to zero valid accelerometer device wearing time [35]. Thus, the total sample size used in this analysis was 3306 corresponding to 3306 EMA prompts from 134 subjects, with between 4 to 35 observations nested within each subject.

Figure 5.1: Distribution of the Outcome Variable: MVPA in Seconds



We hypothesized that the probability of spending > 0 time in MVPA, and the mean and dispersion level of the > 0 MVPA, were all associated with (a) subject-level variables such as age, Body Mass Index (BMI), (b) a day-level variable, e.g., whether the day was a weekend day, and (c) an observation-level variable, e.g., when the assessment happened during the day (morning, afternoon, evening). In other words, for this analysis, X_{ij}, W_{ij} and A_{ij} were the same set of covariates. Using the real-time stamps of each assessment, we defined 'morning' as before 12:00 pm and 'afternoon' as between 12:00 pm and 6:00 pm and 'evening' as after 6:00 pm. With these cutpoints, 13.12% observations happened in the morning, 50% happened in the afternoon, and 36.88% happened in the evening. Also, 35.31% observations happened during the weekend. Prior to the analysis, we standardized BMI using the sample mean (28.96) and standard deviation (6.58) and centered subject's age around the sample mean (41.98 years). We assumed that the effects of age and BMI were linear. For these data, the covariance between location and scale random effects was assumed to be zero as more complex model led to the non-convergence issue.

5.5.2 Findings for the Fixed Effects

Table 5.3 shows the comparison among NB-Mixed, NBLS, H-NB-MIXED and H-NBLS on MATCH data. For the MATCH data, H-NBLS provided the best fit to the data among the four candidate models. The values of Akaike Information Criterion (AIC) for the non-hurdle models were much higher than the hurdle models. As we have described in the simulation section, ignoring the inflated zero counts which can be prevalent in physical activity data seems to be more detrimental than suppressing the random scale effects.

Unlike the hurdle models, the NB-Mixed and NBLS models included the zero outcomes when estimating the mean and dispersion parameters. For the mean function, four candidate models had consistent findings regarding $\beta's$. Older subjects were found to be less active as β_{Age} was negative, yet not significant. The effect of BMI was consistently negatively associated with mean level of MVPA across the four models, which indicates that subjects with higher BMI had lower level of MVPA than subjects with lower BMI, regardless of the inclusion of the zero outcomes or not. MVPA during weekends compared to weekdays displayed almost no difference. However, subjects had higher MVPA levels in the morning and afternoon than in the evening, and subjects tended to exercise mostly in the morning.

In terms of dispersion, the results of the non-hurdle models and hurdle models were different. The hurdle component explained a large part of the issue of over-dispersion. Including the hurdle component, the estimate of $\tau_{Intercept}$ decreased drastically which suggested that zeroes can increase over-dispersion. The positive effects of BMI in the non-hurdle models were

Table 5.3: Model comparisons on the MATCH data of wave 4: comparing NB-mixed, NBLS, H-NB-mixed, H-NBLS

	NB-mixed			**NBLS**			**H-NB-mixed**			**H-NBLS**		
Param	**Est.**	**S.E.**	**P.V.**	**Est.**	**S.E.**	**P.V.**	**Est.**	**S.E.**	**P.V.**	**Est.**	**S.E.**	**P.V.**
$\beta_{Intercept}$	4.13	0.10	<0.01	4.22	0.10	<0.01	4.76	0.06	<0.01	4.83	0.06	<0.01
β_{Age}	-0.01	0.01	0.36.	-0.01	0.01	0.50.	-0.01	0.01	0.34.	-0.01	0.01	0.36
β_{BMI}	-0.17	0.07	0.02	-0.14	0.07	0.05	-0.11	0.05	0.02	-0.11	0.05	0.03
$\beta_{Weekend}$	-0.03	0.10	0.79	0.00	0.09	1.00	0.05	0.05	0.27	0.01	0.05	0.91
$\beta_{Morning}$	0.55	0.15	0.00	0.57	0.14	0.00	0.35	0.08	<0.01	0.33	0.08	<0.01
$\beta_{Afternoon}$	0.39	0.11	0.00	0.41	0.10	0.00	0.23	0.05	<0.01	0.19	0.05	0.00
$\tau_{Intercept}$	2.11	0.05	<0.01	2.19	0.07	<0.01	-0.16	0.08	0.04	-0.06	0.07	0.38
τ_{Age}	0.01	0.00	0.17	0.01	0.01	0.50	0.00	0.01	0.66	0.01	0.01	0.23
τ_{BMI}	0.09	0.03	0.00	0.10	0.06	0.07	-0.05	0.05	0.33	-0.03	0.04	0.42
$\tau_{Weekend}$	-0.01	0.06	0.89	0.02	0.06	0.79	-0.07	0.08	0.39	-0.11	0.08	0.15
$\tau_{Morning}$	-0.34	0.09	0.00	-0.37	0.09	<0.01	0.22	0.12	0.07	0.17	0.11	0.13
$\tau_{Afternoon}$	-0.38	0.06	<0.01	-0.44	0.06	<0.01	-0.14	0.08	0.10	-0.17	0.08	0.03
$\alpha_{Intercept}$							-0.11	0.08	0.18	-0.11	0.08	0.18
α_{Age}							-0.01	0.01	0.54	-0.01	0.01	0.54
α_{BMI}							-0.13	0.06	0.03	-0.13	0.06	0.03
$\alpha_{Weekend}$							-0.01	0.08	0.85	-0.01	0.08	0.85
$\alpha_{Morning}$							0.49	0.12	<0.01	0.49	0.12	<0.01
$\alpha_{Afternoon}$							0.48	0.08	<0.01	0.48	0.08	<0.01
$ln(\sigma_{\nu}^2)$	-1.28	0.25	<0.01	-1.64	0.28	<0.01	-1.54	0.17	<0.01	-1.52	0.16	<0.01
$ln(\sigma_{\omega}^2)$				-1.21	0.18	<0.01				-2.08	0.27	<0.01
$ln(\sigma_{\psi}^2)$							-1.12	0.20	<0.01	-1.12	0.20	<0.01
-2 log lik	28027			27856			26489			26448		
AIC	28053			27884			26529			26490		

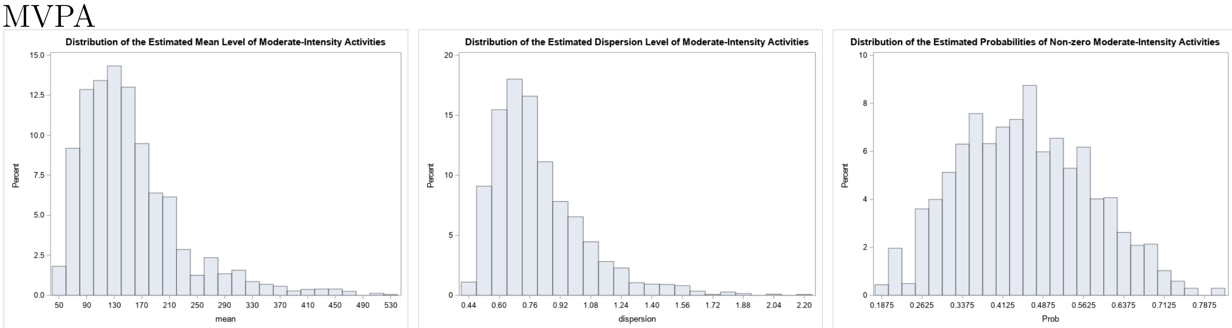
Param: Parameters
Est.: estimate
S.E.: standard errors
P.V.: p-value
-2 log lik: -2 log likelihood value
AIC: Akaike Information Criterion

significant or marginally significant but after we included the hurdle, it became insignificant and the direction of the effect even changed. In the hurdle candidate models, the effect of BMI on dispersion became negative, though insignificant, which suggests that higher BMI was associated with lower dispersion levels among the positive valued outcomes. In other words, subjects with higher BMI tended to exhibit greater consistency in the amount of MVPA than subjects with lower BMI when they had MVPA. Still the dispersion level was not significantly different between weekends and weekdays. However, the dispersion level was significantly different across morning through evening. In the NB-Mixed and NBLS models, where zeros were included to estimate the mean and dispersion, the dispersion level of MVPA was higher in the evening than in the morning and afternoon. However, if we only included the positive valued outcomes which is the case of hurdle models, then morning has the highest level of dispersion.

In terms of the probability of having non-zero MVPA, BMI had a significant effect such that higher BMI levels were associated with a lower probability of MVPA. For the time of day indicators, the probability of MVPA was increased in the morning and afternoon. According to the descriptive analysis, 60.18% observations in the morning had positive MVPA, while only 59.23% and 48.19% observations exhibited positive MVPA in the afternoon and evening.

5.5.3 Interpretation of Random Effects

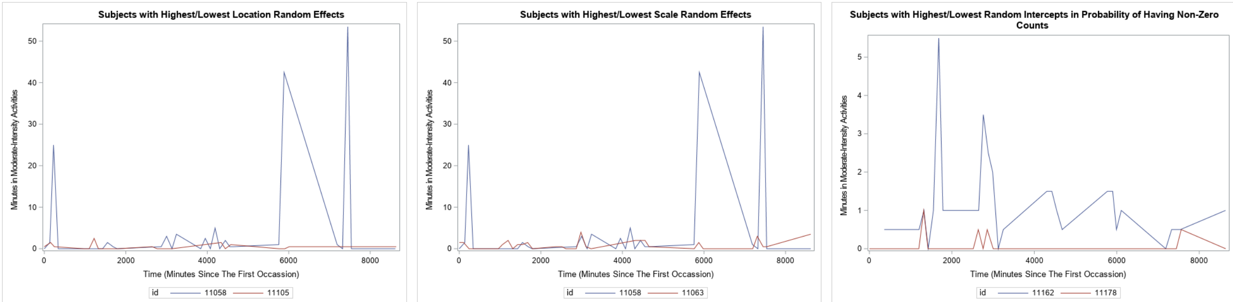
Figure 5.2: Distribution of Estimated Mean, Dispersion and Probability of Having Positive MVPA



The random location and scale effects summarize the effect of unobserved subject-level variables associated with the subject's mean and dispersion, while the random intercept in the hurdle component summarizes the subject-level unobserved effect related to the probability of having non-zero MVPA. Figure 5.2 shows the density of estimated mean, dis- persion and probability of having non-zero MVPA which were calculated using the estimates of $\beta, \tau, \alpha, \nu_i, \omega_i, \psi_i$ from H-NBLS, the best-fitting model. The empirical bayes estimates of ν_i ranged from -0.93 to 1.08, ω_i estimates ranged from -0.37 to 0.74, and ψ_i estimates ranged from -1.31 to 1.04, so that the resulting mean, dispersion, and probability ranged from 46.00$\sim$ 536.21 seconds, 0.42$\sim$2.22 and 0.19$\sim$0.73, respectively.

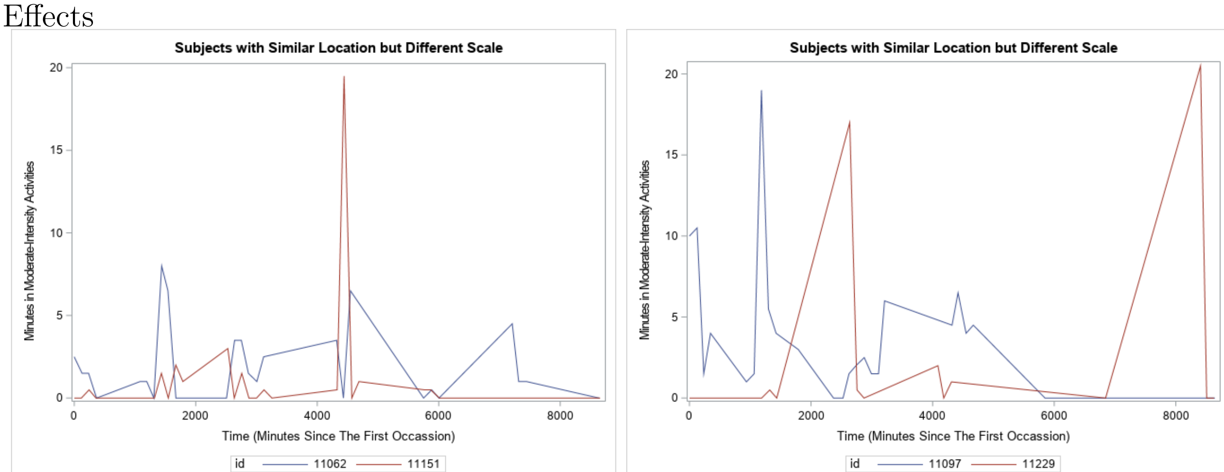
From Figure 5.3, Subject 11058 was identified as having the most positive location and scale

Figure 5.3: Subjects with Highest/Lowest Random Effect Estimates



random effect estimates, Subject 11105 had the most negative location random effect estimate, and Subject 11063 had the most negative scale random effect estimate. We can see that Subject 11058 had high average MVPA amount but the amount drastically changed over time and was highly dispersed. Subject 11105 and 11063 displayed a flat curve with MVPA amounts close to zero. By this comparison, we can see that ν_i, ω_i were consistent with what we have observed in MVPA level. ψ_i is the random effect related to the frequency that subjects have positive MVPA, or in a more intuitive way, the frequency that subjects are active. We can see that the MVPA amount of Subject 11178 is mostly zero across all observations, while the MVPA of Subject 11162 was zero only 3 times and instead the subject was active at most of the observations.

Figure 5.4: Subjects with Similar Location Random Effects but Different Scale Random Effects



From the results above, the random location effect mainly reflects the average positive MVPA level of the subject when they exercise. However, even for subjects with similar location random effects, they can have different dispersion levels (Figure 5.4). For example, Subject 11151 and 11062 had similar values (0.049 vs 0.051) for location random effects, but Subject 11151 was more dispersed than Subject 11062 as they had different scale random effect estimates (0.293 vs. -0.180). Subject 11151 had a very high MVPA level at one time point, but for other times the subject's MVPA levels were almost zero. Conversely, Subject 11062's MVPA levels were more consistent. As another example, Subject 11229 and 11197 had location and scale levels respectively as (0.662 vs. 0.659) and (0.407 vs. -0.234), and Subject 11229's MVPA levels were more dispersed over time. Also, for subjects with similar probabilities of having positive MVPA, they can be different in terms of the mean and dispersion of MVPA. This provides some justification for our proposed model that allows estimation of each subject's mean, dispersion, and the probability of non-zeros for physical activity data.

5.6 Discussion

In this article, we proposed a negative binomial mixed effect location-scale model for intensive longitudinal count data. This model provides random effects which characterize subjects' exercise patterns in three dimensions: the mean level, the dispersion level, and the probability of having non-zero MVPA. Such information may be useful for subject-level classification and for prediction. In this article, we displayed some graphical descriptions for certain selected subjects. However, classification and clustering algorithms could also be used to divide the subjects into informative groups, and perhaps one can examine how subjects from these groups maintain long-term outcomes such as 12-month recommended MVPA.

In this study, we used maximum likelihood estimation with Adaptive Gauss-Hermite Quadrature to approximate the marginal likelihood. The computational cost for the application example was reasonable as it took 15-20 minutes to converge for the dataset with 134 subjects and 4 to 35 observations within each subject. We also tried to fit a three-level model (e.g., subject-level, day-level and observation-level) and specified day-to-day scale random effects, but it took much longer to converge as the complexity of this approximation algorithm grows exponentially with the number of random effects. Other than the maximum likelihood approach, using alternative estimation methods or Bayesian sampling can potentially provide more efficient solutions for estimating more complex multi-level mixed-effect location-scale models.

For this analysis, observations were assumed to be conditionally independent given the subject-level random effects. In the scenario of physical activities, sequential information and the ordering may also be useful. For example, the cumulative amount of MVPA of the past few days might be associated with subject's decision to do MVPA today. Future work can also focus on the utilization of the sequential information and even the sequential prediction for the amount of physical exercise on each day.

Finally we should note that, as we used the negative binomial distribution to model these count data, we assumed every subject exhibited some degree of over-dispersion. According to the probability mass function of the negative binomial model, i.e., $\sigma^2 = \mu + \frac{1}{r}\mu^2$, $\frac{1}{r}$ is a positive value which can only approximate to zero but can never be exactly zero. Thus, the proposed model is less sufficient to handle count data which are under-dispersed or non-dispersed.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Summary

In this dissertation, we proposed several statistical models motivated by specific issues of intensive longitudinal data collected by mobile or wearable devices. And all of these methods focus on inferring the latent effects at subject-level or even a day-level to account for the heterogeneity in mean and even in variability.

For the first work, the three-level mixed effect model with separate autocorrelated random effects at between-day level and within-day level characterized the correlation across day and also within day and the extent of correlation was exponentially related to the interval length between consecutive assessments.

For the second work, a convenient joint modelling approach for the response process and the measurement process was proposed and the inter-correlation between them was described in terms of latent effects, which potential unveiled the mechanism of non-ignorable nonresponses in EMA studies. Subject-level latent effects associated with the mean and variability of the event-contingent assessments can be correlated with the responsivity of the event reports. We found that, for each subject the improvement of mood as well as higher variability of mood were potentially associated with the responsivity. And via this model we can also compare the same fixed effects on the outcome mean, the outcome variability as well as the responsivity. And this method is one of the first few works that address the issue of responsivity of the self-initiated event-contingent EMA assessments.

For the third work, we framed the problem of identical responses as the extreme random scale effects in the variability distribution and therefore latent clustering can be used to classify subjects with varying responses and identical responses. And this is one of the first works which deal with the duplicate data in longitudinal surveys. And we will also discuss the similarity in terms of the data generating mechanism between these duplicate data and missing data.

The last work is a model which deals with the physical activity data collected from wearable devices. Therefore, the estimated location random effects and dispersion effects were informative to represent some the subject-level traits in subjects' physical activity patterns. The proposed model is one of the first few works to deal with physical activity data collected wearable devices.

6.2 Potential Future Directions

The past ten years have witnessed a huge progress in EMA study design. The involvement of more advanced wearable devices such as wearable cameras and smart watches or even the home sensors have enabled more objective and comprehensive assessments of subjects' real-life experiences and the surrounding environment. An EMA system may include multiple data sources and data from each venue may be of different data-types. In this thesis, we proposed methods for analyzing smartphone-based EMA surveys and accelerometer data which described subjects' psychological states and physical activities in their daily lives. These data and also the physiological data including the heart beats and metabolism rates are often numerical and can be easily quantified. But nowadays EMA studies designs especially those incorporating the image and audio technology may even have data including but not limited to location information, audio recording, real-time images and videos. And these more abstract and complex data necessitates the use of artificial intelligence techniques for image segmentation, object detection and environment classification.

There might be several interesting topics in mHealth data analysis await for statisticians as well as data scientists to explore. The first one is to reasonably and effectively combine these complex data from multiple data sources with the constraint of limited computational resources. More specifically, unlike conventional longitudinal data which are equally spaced and balanced, these EMA data are recorded in real time and the number of repeated measures for each outcome may probably be different, which calls for methods or norms for data alignment and reconstruction.

And beyond this, methods for modelling multiple outcomes from a more systematic scope rather than treating them separately can be useful to reveal the intercorrelation network among these measures. Majority parts of this thesis are for single outcome modelling. But rather than to just focus on one particular measure, it might be more informative to model a comprehensive set of multiple outcomes. Methods proposed in Chapter 3 was an attempt to summarize the bivariate/multivariate dependence among outcomes via the shared random effects. But as we have discussed in the end of Chapter 3, methods that are able to directly model multiple outcomes are warranted.

The second one might be the potential high-dimensional and sparse issues. As we have mentioned above, the after processed data from audio or video can be subject to these issues. These data might probably be categorical instead of numerical, which means those can not be fully quantified. Thus, modern machine learning algorithms as well as the high-dimensional statistical methods might be necessary for analyzing such data. But even for these novel types of data, the fact we can't ignore is that these data are hierarchically clustered, longitudinal and intensive. However, most of the existing high-dimensional methods and machine learning algorithms are not applicable for complex large-scale multilevel data. Future studies can focus on evolving these methods to their multilevel longitudinal versions.

In addition, the method proposed in this thesis assumed these repeated measures to be conditionally independent. However, by doing this, the sequential information was ignored. EMA assessments are intensive so that assessments might be highly correlated. Methods utilizing the sequential correlation of intensive longitudinal data should be proposed to improve the forecasting accuracy. Moreover, the convenient intensive data collection approach can probably result in high-volume data flux so that it may be expensive to store the large-scale historical data. Conventional analysis framework where model were trained and estimated using historical datasets might be computationally intensive. Instead, current methods should also be extended to enable data streaming analyses and prediction.

REFERENCES

[1] Alan Agresti, Brian C. Caffo, and Pamela Ohman-Strickland. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Journal of Computational Statistics and Data Analysis*, 47:2088–2112, 2007.

[2] Paul Allison. Don't put lagged dependent variables in mixed models, 2019. URL `https://statisticalhorizons.com/lagged-dependent-variables`.

[3] Dickson Anumendem, Geert Verbeke, Bieke De Fraine, Patrick Onghena, and Jan Van Damme. Double serial correlation for multilevel growth curve models. *Quality & Quantity: International Journal of Methodology*, 47(3):1413–1427, April 2013.

[4] S. Baldwin, G. Fellingham, and Austin S Baldwin. Statistical models for multilevel skewed physical activity data in health research and behavioral medicine. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, 35 6:552–62, 2016.

[5] R Darrell Bock and Stephen Henry Charles du Toit. Parameter estimation in the context of non-linear longitudinal growth models. In *Methods in Human Growth Research*, pages 234–257. Cambridge University Press, 2004.

[6] Niall Bolger and Jean-Philippe Laurenceau. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*, chapter Limitation of Intensive Longitudinal Designs, page 21. Guilford Press, 2007.

[7] Dena M. Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison L Gienger, Nancy D Lin, Robyn Lewis, Christopher Stave, Ingram Olkin, and John R Sirard. Using pedometers to increase physical activity and improve health: A systematic review. *Journal of American Medicine Association*, 298:2296–2304, 2007.

[8] Meredith A. Case, Holland A. Burwick, Kevin G. Volpp, and Mitesh S. Patel. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*, 313:625–626, 2015.

[9] Junliang Chen, Daowen Zhang, and Marie Davidian. A monte carlo em algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, 3:347–360, 2002.

[10] Eric M. Chi and Gregory C. Reinsel. Models for longitudinal data with random effects and ar(1) errors. *Journal of the American Statistical Association*, 84(406):452–459, 1989.

[11] Stephen S Coughlin and Jessica L. Stewart. Use of consumer wearable devices to promote physical activity: A review of health intervention studies. *Journal of Environment and Health Science.*, 2:2378–6841, 2016.

[12] An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G. Kenward. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Journal of Educational and Behavioral Statistics*, 52:111–125, 2010.

[13] An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G. Kenward. Generalized shared-parameter models and missingness at random. *Journal of Educational and Behavioral Statistics*, 11:279–310, 2011.

[14] John F Cursio, Robin J Mermelstein, and Donald Hedeker. Latent trait shared-parameter mixed models for missing ecological momentary assessment data. *Statistics in Medicine*, 38:660–673, 2019.

[15] Arthur P Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1): 1–38, 1977.

[16] Peter J. Diggle. An approach to the analysis of repeated measurements. *Biometrics*, 44 (4):959–971, 1988.

[17] Genevieve F. Dunton, Yue Liao, Eldin Dzubur, Adam M. Leventhal, Jimi Huh, Tara Gruenewald, Gayla Margolin, Carol Koprowski, Eleanor Tate, and Stephen Intille. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: Protocol for the match study. *Contemporary Clinical Trials*, 43:142–154, 2015.

[18] Silvia Fruhwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer Science+Business Media, LLC, New York, 2006.

[19] Sujuan Gao. A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Statistics in Medicine*, 23:213–219, 2004.

[20] Weihua Gao, Donald Hedeker, Robin J. Mermelstein, and Hui Xie. A scalable approach to measuring the impact of nonignorable nonresponse with an ema application. *Statistics in Medicine*, 35:5579–5602, 2016.

[21] Nisha C. Gottfredson, Daniel J. Bauer, and Scott A. Baldwin. Modeling change in the presence of nonrandomly missing data: Evaluating a shared parameter mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2):196–209, 2014.

[22] Jeffrey R. Harring and Shelley A. Blozis. Fitting correlated residual error structures in nonlinear mixed-effects models using sas proc nlmixed. *Behavior Research Methods*, 46: 372–384, 2014.

[23] Donald Hedeker, Robin J. Mermelstein, and Hakan Demirtas. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (ema) data. *Biometrics*, 64:627–634, 2008.

[24] Donald Hedeker, Hakan Demirtas, and Robin J. Mermelstein Mermelstein. A mixed ordinal location scale model for analysis of ecological momentary assessment (ema) data. *Statistics and its interface*, 2:391–401, 2009.

[25] Donald Hedeker, Robin J. Mermelstein, and Hakan Demirtas. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, 31, 2012. doi: 10.1002/sim.5338.

[26] Michael R. Hufford and Saul Shiffman. Assessment methods for patient-reported outcomes. *Disease Management and Health Outcomes*, 11:77–86, 2003.

[27] Michael P. Jones. Indicator and stratification methods for missing explanatory variables in multiple linear regression, 1996.

[28] Stephanie A. Kovalchik, Steven C. Martino, Rebecca L. Collins, William G. Shadel, Elizabeth J. D'Amico, and Kirsten Becker. Scaled inverse probability weighting: A method to assess potential bias due to event nonreporting in ecological momentary assessment studies. *Journal of Educational and Behavioral Statistics*, 43:354–381, 2018.

[29] Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[30] Kenneth Lange. *Numerical Analysis for Statisticians.* Springer-Verlag New York, New York, 1999.

[31] Xiaolei Lin, Robin Mermelstein, and Donald Hedeker. A shared parameter location scale mixed effect model for ema data subject to informative missing. *Health Services and Outcomes Research Methodology*, 18:227–243, 2018.

[32] Saskia Litiere, Ariel Alonso, and Geert Molenberghs. The impact of a misspecified random-effects distributionon the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 27:3125–3144, 2008.

[33] Roderick J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121, 1995.

[34] Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data.* John Wiley and Son, Inc., Hoboken, New Jersey, 2019. ISBN 978-0-470-52679-8.

[35] Jaclyn P. Maher, Eldin Dzubur, Rachel Nordgren, Jimi Huh, Chih-Ping Chou, Donald Hedeker, and Genevieve F. Dunton. Do fluctuations in positive affective and physical feeling states predict physical activity and sedentary time? *Psychology of Sport and Exercise*, 41:153–161, 2019. ISSN 1469-0292. doi: https://doi.org/10.1016/j.psychsport.2018.01.011. URL `https://www.sciencedirect.com/science/article/pii/S1469029217305691`.

[36] Marianthi Markatou. Mixture models, robustness, and the weighted likelihood methodology. *Biometrics*, 56(2):483–486, 2000. doi: https://doi.org/10.1111/j.0006-341X.2000.00483.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2000.00483.x.

[37] Geoffrey J. McLachlan. *Finite Mixture Models*. Wiley, 2000.

[38] AW Meade and SB Craig. Identifying careless responses in survey data. *Psychological Methods*, 17(3):235–394, 2012.

[39] Robin J. Mermelstein and Kathleen R. Diviak. Context and subjective experience surrounding dual cigarette and e-cigarette use, 2019 (accessed: 11/05/2019). URL https://www.ihrp.uic.edu/study/context-and-subjective-experience-surrounding-dual-cigarette-and-e-cigarette-use.

[40] Yongyi Min and Alan Agresti. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5:1–19, 2005.

[41] Debbie S. Moskowitz and Simon N. Young. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, 31(1):13–20, 2006.

[42] David I. Ohlssen, Linda D. Sharples, and David J. Spiegelhalter. Flexible random-effects models using bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*, 26:2088–2112, 2007.

[43] Hans-Peter Piepho and Charles E. McCulloch. Transformations in mixed models: Application to risk analysis for a multienvironment trial. *Journal of Agricultural, Biological, and Environmental Statistics*, 9:123–137, 2004.

[44] JosÃ© C Pinheiro and Edward C Chao. Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1):58–81, 2006.

[45] Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21, 2002. doi: 10.1177/1536867X0200200101. URL https://doi.org/10.1177/1536867X0200200101.

[46] Richard A. Redner and Homer F. Walker. Mixture densities maximum likelihood and the em algorithm. *SIAM Review*, 26:195–239, 1984.

[47] Jason Roy. Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical Methods in Medical Research*, 16:441–456, 2007.

[48] Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Son, Inc., Hoboken, New Jersey, 2004. ISBN 978-0-471-08705-2.

[49] Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22:278–295, 2011.

[50] Lewis B. Sheiner and Stuart L. Beal. Evaluation of methods for estimating population pharmacokinetic parameters. iii. monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 11:303–319, 1983.

[51] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4:1–32, 2008.

[52] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227 – 244, 2000. ISSN 0378-3758. doi: https://doi.org/10.1016/S0378-3758(00)00115-4. URL `http://www.sciencedirect.com/science/article/pii/S0378375800001154`.

[53] Judith D. Singer. Using sas proc mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23: 323–355, 2018.

[54] Alexander W. Sokolovsky, Robin J. Mermelstein, and Donald Hedeker. Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine and Tobacco Research*, 16:351–358, 2014.

[55] Arthur A. Stone, Saul Shiffman, Audie A. Atienza, and Linda Nebeling. *The Science of Real-Time Data Capture: Self-Reports in Health Research*, chapter Historical Roots and Rationale of Ecological Momentary Assessment (EMA), pages 3–10. Oxford University Press, 2007.

[56] Roula Tsonaka, Geert Verbeke, and Emmanuel Lesaffre. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, 65: 81–87, 2009.

[57] Kristof Vansteelandt and Geert Verbeke. A mixed model to disentangle variance and serial autocorrelation in affective instability using ecological momentary assessment data. *Multivariate Behavioral Research*, 51(4):446–465, 2016.

[58] Geert Verbeke and Emmanuell Lesaffre. A linear mixed-effects model with heterogeneity in the random-reffects population. *Journal of the American Statistical Association*, 91: 217–221, 1996.

[59] Geert Verbeke and Geert Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer-Verlag New York, New York, 2000.

[60] Reza Drikvandiand Geert Verbeke and Geert Molenberghs. Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*, 73:63–71, 2017.

[61] Matthew Willetts, Sven Hollowell, Louis Aslett, Chris Holmes, and Aiden Doherty. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 uk biobank participants. *Scientific Report*, 8:7961, 2018.

[62] Weichun Xu and Donald Hedeker. A random-effects mixture model for classifying treatment response in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, 11:253–273, 2001.

[63] Kelvin K. W. Yau, Kui Wang, and Andy H. Lee. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biom. J.*, 45:437–452, 2003.

[64] Ashenafi A. Yirga, Sileshi F. Melesse, Henry G. Mwambi, and Dawit G. Ayele. Negative binomial mixed models for analyzing longitudinal cd4 count data. *Scientific Report*, 10: 16742, 2020.

[65] Ying Yuan and Guosheng Yin. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Statistical Methods & Applications*, 66:105–114, 2010.

[66] Peng Zhang, Peter X. K. Song, Annie Qu, and Tom Greene. Efficient estimation for patient-specific rates of disease progression using nonnormal linear mixed models. *Biometrics*, 64:29–38, 2008.

[67] Xinyan Zhang, Himel Mallick, Zaixiang Tang, Lei Zhang, Xiangqin Cui, Andrew K. Benson, and Nengjun Yi. Negative binomial mixed models for analyzing microbiome count data. *BMC Bioinformatics*, 8:4, 2017.