Health Services and Outcomes Research Methodology
https://doi.org/10.1007/s10742-020-00220-w

A three-level mixed model to account for the correlation at both the between-day and the within-day level for ecological momentary assessments

Qianheng Ma[1] · **Robin Mermelstein**[2] · **Donald Hedeker**[1]

Received: 5 May 2020 / Revised: 17 August 2020 / Accepted: 5 September 2020

Abstract

Ecological Momentary Assessment (EMA) studies aim to explore the interaction between subjects' psychological states and real environmental factors. During the EMA studies, participants 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). Those three-level data may exhibit complex longitudinal correlation structure but ignoring or mis-specifying the within-subject correlation structure can lead to bias on the estimation of the key effects and the intraclass correlation. Given the three-level EMA data and the time stamps of the responses, we proposed a linear mixed effects model with random effects at each level. In this model, we accounted for level-2 autocorrelation and level-1 autocorrelation and showed how structural information from the three-level data improved the fit of the model. With real time stamps of the assessments, we also provided a useful extension of this proposed model to deal with the issue of irregular-spacing in EMA assessments.

Keywords EMA · Intensive multilevel longitudinal data · Mixed effect model · AR(1) · Irregular-spaced assessments

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10742-020-00220-w) contains supplementary material, which is available to authorized users.

✉ Qianheng Ma
qma@uchicago.edu

Robin Mermelstein
robinm@uic.edu

Donald Hedeker
Dhedeker@health.bsd.uchicago.edu

[1] Department of Public Health Sciences, University of Chicago, Chicago, USA

[2] Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, USA

Published online: 23 September 2020

1 Introduction

The Ecological Momentary Assessment (EMA) design (Shiffman et al. 2008) 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 (Sokolovsky et al. 2013) 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.

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

Mixed effect models (Laird and Ware 1982) 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 (1989) 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 (Eq. (1)).

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

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 (Verbeke and Molenberghs 2000). Equation (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.

$$\mathrm{P}_{SAR(1)} = \begin{bmatrix} 1 & \rho^{|t_2-t_1|} & \dots & \rho^{|t_{n_i}-t_1|} \\ \rho^{|t_1-t_2|} & 1 & \dots & \rho^{|t_{n_i}-t_2|} \\ \vdots & \vdots & & \vdots \\ \rho^{|t_1-t_{n_i}|} & \rho^{|t_2-t_{n_i}|} & \dots & 1 \end{bmatrix} \tag{2}$$

While Chi and Reinsel (1989) 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 (Diggle 1988). This model with this decomposition is proved to be a more general form than Chi and Reinsel's model (Verbeke and Molenberghs 2000). 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 (Diggle 1988), sometimes the data may not contain sufficient information to estimate both the subject random effects v_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.

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 v_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 (Vansteelandt and Verbeke 2016). 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. (2013) 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.

1.3 Include lagged dependent variable as model covariates

Another intuitive way to account for the autocorrelation is to include the lagged dependent variable (Allison 2019) 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 (Allison 2019) 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 The proposed method

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 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 (AR(1)$_d$) and AR(1) correlated within-day assessment errors (AR(1)$_{wd}$). Sections 2.2.1 to 2.2.3 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.

2.2.1 Three-level mixed effect model

For a three-level mixed effects model (Eq. (3)), there is an independent random subject-level intercept v_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. v_i, ψ_{ij} and ϵ_{ijk} all follow zero-mean normal distributions but with different covariance structures. The variance of v_i is σ_v^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.

$$y_{ijk} = X_{ijk}^T \beta + v_i + \psi_{ij} + \epsilon_{ijk} \tag{3}$$

2.2.2 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 (Eq. (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 (Eq. (2)) if time-stamped information is provided.

$$\begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij1} \\ \vdots \\ \epsilon_{ijN_{ij}} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1^{N_{ij}-1} \\ \rho_1 & 1 & \dots & \rho_1^{N_{ij}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{N_{ij}-1} & \rho_1^{N_{ij}-2} & \dots & 1 \end{pmatrix} \sigma_\epsilon^2 \right] \tag{4}$$

2.2.3 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 (Eq. (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.

$$\begin{pmatrix} \psi_{i1} \\ \psi_{i2} \\ \vdots \\ \psi_{iN_i} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2 & \dots & \rho_2^{N_i-1} \\ \rho_2 & 1 & \dots & \rho_2^{N_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_2^{N_i-1} & \rho_2^{N_i-2} & \dots & 1 \end{pmatrix} \sigma_\psi^2 \right] \tag{5}$$

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.

3 Simulation setting

3.1 Model comparison

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

3.1.1 Model A: RI$_s$+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 (AR(1)$_{ws}$) (Eq. (6)).

$$y_{ik'} = X_{ik'}^T \beta + v_i + \epsilon_{ik'} \tag{6}$$

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 (Eq. (7)) for each subject i .

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i\Sigma_j N_{ij}} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \dots & \rho^{\Sigma_j N_{ij}-1} \\ \rho & 1 & \dots & \rho^{\Sigma_j N_{ij}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{\Sigma_j N_{ij}-1} & \rho^{\Sigma_j N_{ij}-2} & \dots & 1 \end{pmatrix} \sigma^2_{\epsilon'}\right] \tag{7}$$

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

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

Model B is a three-level mixed effects model (Eq. (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 Eq. (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 (Eq. (8)).

$$\begin{pmatrix} \psi_{i1} \\ \psi_{i2} \\ \vdots \\ \psi_{iN_i} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, I_{N_i}\sigma^2_{\psi}\right] \tag{8}$$

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

3.1.3 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 (Eq. (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.

$$y_{ijk} = X^T_{ijk}\beta + v_i + \psi_i jk + \rho_1 y_{ij(k-1)} + \rho_2\overline{y_{i(j-1)}} + \epsilon_{ijk} \tag{9}$$

3.2 Data generation

For true values of the model parameters, for variances, we set $\sigma^2_v = exp(1)$ and $\sigma^2_\psi = exp(0.5)$ and $\sigma^2_\epsilon = exp(0.5)$ 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 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.

Table 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+ $AR(1)_d$+ $AR(1)_{wd}$		
	CR	Mean(SD)	CR	Mean(SD)	CR	Mean(SD)	CR	Mean(SD)	True Value
β_{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
β_{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
β_{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
β_{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
$log(\sigma^2_v)$	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
$log(\sigma^2_\psi)$			0.881	0.400(0.101)	0.920	0.466(0.077)	0.952	0.497(0.108)	0.5
$log(\sigma^2_\epsilon)$	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
$logit(\rho_1)$	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
$logit(\rho_2)$					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

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 v_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 Eq. (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 Eq. (4).
5. Then we simulated the outcome variable Y_{ijk} according to Eq. (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:

$$Coverage = \frac{\text{number of times that the 95\% CIs covers the true value}}{\text{number of successfully convergent solutions}} \times 100\% \quad (10)$$

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 Sect. 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.

4 Estimation

4.1 Objective functions for optimization

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

According to Sect. 3.1.1, given the v_i, the conditional likelihood for the random error vector $\epsilon_i' = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{i(\Sigma_j N_{ij})})^T$ is,

$$f(\epsilon_i'|v_i) = \frac{1}{|2\pi\Sigma_{\epsilon'}|^{1/2}} exp(-\frac{1}{2}\epsilon_i'^T \Sigma_{\epsilon_i'} \epsilon_i') \tag{11}$$

The form of the variance-covariance matrix Σ_ϵ' is shown in Eq. (7). For estimating the extended version of this model, we replaced the regular AR(1) structure by serial AR(1) structure as shown in Eq. (2).

The distribution of v_i (a scalar) is,

$$g_3(v_i) = \frac{1}{\sqrt{(2\pi\sigma_v^2)}} exp(-\frac{v_i^2}{2\sigma_v^2}) \tag{12}$$

So the marginal likelihood for ϵ_i' is,

$$\mathcal{L}(\epsilon') = \int \prod_{i=1}^{N_i} \{f(\epsilon_i'|v_i) g_3(v_i) dv_i\} \tag{13}$$

4.1.2 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 $\epsilon_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \dots, \epsilon_{ijN_{ij}})^T$.

$$f(\epsilon_{ij}|\nu_i, \psi_{ij}) = \frac{1}{|2\pi \Sigma_\epsilon|^{1/2}} exp(-\frac{1}{2}\epsilon_{ij}^T \Sigma_\epsilon \epsilon_{ij}) \tag{14}$$

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

$$g_2(\psi_i) = \prod_{j=1} \frac{1}{\sqrt{2\pi\sigma_\psi^2}} exp(-\frac{\psi_{ij}^2}{2\sigma_\psi^2}) \tag{15}$$

The distribution of ν_i (a scalar) is,

$$g_3(\nu_i) = \frac{1}{\sqrt{(2\pi\sigma_\nu^2)}} exp(-\frac{\nu_i^2}{2\sigma_\nu^2}). \tag{16}$$

So the marginal likelihood for optimization is given as,

$$\mathcal{L}(\epsilon) = \int_\nu \int_\psi \prod_{i=1}^{N} \{ \prod_{j=1}^{N_i} \{ f(\epsilon_{ij}|\nu_i, \psi_{ij}) g_2(\psi_{ij}) d\psi_{ij} \} g_3(\nu_i) d\nu_i \}. \tag{17}$$

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.

4.1.3 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 (1996) to prevent these observations to be deleted during programming so that we could still use all the available data.

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

Given ψ_{ij} and ν_i, the conditional likelihood for the random error vector will still follow Eq. (14) but the variance-covariance matrix Σ_ϵ will be in the AR(1) form (Eq. (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 $\psi_i = (\psi_{i1}, \psi_{i2}, \dots, \psi_{iN_i})^T$is,

$$g_2(\psi_i) = \frac{1}{|2\pi \Sigma_\psi|^{1/2}} exp(-\frac{1}{2}\psi_i^T \Sigma_\psi \psi_i) \tag{18}$$

And the variance-covariance matrix of the day-level random effects is specified by AR(1) structure (Eq. (5)).

The distribution of v_i (a scalar) is the same as Eq. (12). So the objective function for optimization is given as,

$$\mathcal{L}(\epsilon) = \int_{v} \int_{\psi} \prod_{i=1}^{N} \{ f(\epsilon_i | v_i, \psi_i) g_2(\psi_i) g_3(v_i) d\psi_i dv_i \} \tag{19}$$

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 (Pinheiro and Chao 2006) 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.

5 Results of simulation study

5.1 Comparing different model structures

Table 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 4.1.4 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 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.

5.2 Comparing proposed models with AR(1) versus with serial AR(1) structure

Table 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

Table 2 Model Comparison over 500 datasets simulated from the extended proposed model with serial AR(1) errors

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

RI_s: independent subject-level random intercept(RI).

$AR(1)_d$: AR(1) correlated random day effects (level-2).

$SAR(1)_{wd}$: Serial-AR(1) correlated within-day(wd) random errors.

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 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 and variances can be problematic.

6 Application on adolescent smoking study data

6.1 The adolescent smoking study data

There were 461 subjects in the Adolescent Smoking Study (Sokolovsky et al. 2013). 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 and 2.2.3. 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.

Table 3 Model comparison over 500 datasets simulated from the original proposed model with serial AR(1) errors

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

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 adjusted the day ID according to the time stamps: we assumed the first 3 h 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 Sect. 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-centered *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.

Table 4 Model comparison on data from Adolescent Smoking Study (original versions)

Parameters	Model A RI_s+AR(1)$_{ws}$	Model B RI_s+RI_d+AR(1)$_{wd}$	Proposed Model RI_s+AR(1)$_d$+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
−2*loglik*	47083	47044	46883
AIC	47105	47068	46909

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

Table 5 Model comparison on data from Adolescent Smoking Study (extended versions)

Parameters	Model A RI_s+$AR(1)_{ws}$	Model B RI_s+RI_d+$SAR(1)_{wd}$	Proposed Model RI_s+$AR(1)_d$+$SAR(1)_{wd}$
$\beta_{Intercept}$	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_v	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;

6.2 Results

From Tables 4 and 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 ($\sigma^2_v = 1.685$). As we introduced more sources of variation, the subject variance displayed a decreasing trend from Model A to Model B ($\sigma^2_v = 1.673$) to our proposed model ($\sigma^2_v = 0.747$). Especially after we accounted for the day-to-day correlation, σ^2_v 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 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_v^2'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.

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 Sect. 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 (2014) 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 (Sheiner and Beal 1983). 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.

Funding The project described in this article (the Adolescent Smoking Study) was supported by a grant from the National Cancer Institute (P01 CA098262). And this work was supported by a grant from the National Cancer Institute (R01 CA240713).

Compliance with ethical standard

Conflict of interest The content is solely the responsibility of the authors and does not necessarily represent the official views of National Cancer Institute.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Allison, P.:Dont put lagged dependent variables in mixed models. (2019). https://statisticalhorizons.com/lagged-dependent-variables

Anumendem, D., Verbeke, G., Fraine, B.D., Onghena, P., Damme, J.V.: Double serial correlation for multilevel growth curve models. Qual. Quant. Int. J. Methodol. **47**(3), 1413–1427 (2013)

Chi, E.M., Reinsel, G.C.: Models for longitudinal data with random effects and ar(1) errors. J. Am. Stat. Assoc. **84**(406), 452–459 (1989)

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

Harring, J.R., Blozis, S.A.: Fitting correlated residual error structures in nonlinear mixed-effects models using sas proc nlmixed. Behav. Res. Methods **46**, 372–384 (2014)

Jones, M.P.: Indicator and stratification methods for missing explanatory variables in multiple linear regression. J. Am. Stat. Assoc. **91**(433), 222–230 (1996)

Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. Biometrics **38**(4), 963–974 (1982)

Pinheiro, J.C., Chao, E.C.: Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. J. Comput. Graph. Stat. **15**(1), 58–81 (2006)

Sheiner, L.B., Beal, S.L.: Evaluation of methods for estimating population pharmacokinetic parameters. iii. monoexponential model: Routine clinical pharmacokinetic data. J. Pharmacokinet. Biopharm. **11**, 303–319 (1983)

Shiffman, S., Stone, A.A., Hufford, M.R.: Ecological momentary assessment. Ann. Rev. Clin. Psychol. **4**(1), 1–32 (2008)

Sokolovsky, A.W., Mermelstein, R.J., Hedeker, D.: Factors predicting compliance to ecological momentary assessment among adolescent smokers. Nicotine Tob. Res. **16**(3), 351–358 (2013)

Vansteelandt, K., Verbeke, G.: A mixed model to disentangle variance and serial autocorrelation in affective instability using ecological momentary assessment data. Multivar. Behav. Res. **51**(4), 446–465 (2016)

Verbeke, G., Molenberghs, G.: Linear Mixed Models for Longitudinal Data. Springer, New York (2000)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.