Title

A model for repeated clustered data with informative cluster sizes

Journal

Statistics in Medicine, 33(5)

ISSN

0277-6715

Authors

Iosif, Ana-Maria

Sampson, Allan R

Publication Date

2014-02-28

DOI

10.1002/sim.5988

Peer reviewed

Published in final edited form as:
Stat Med. 2014 February 28; 33(5): 738–759. doi:10.1002/sim.5988.

A model for repeated clustered data with informative cluster sizes

Ana-Maria Iosif[a,*,†] and **Allan R. Sampson**[b]

[a]Division of Biostatistics, Department of Public Health Sciences, University of California Davis, Davis, CA 95616, U.S.A

[b]Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A

Abstract

Many chronic diseases or health conditions manifest with recurring episodes, each of which can be characterized by a measure of intensity or severity. Both the number of episodes and the severity of each episode can depend on the latent severity of an individual's underlying condition. Data such as this are commonly gathered repeatedly at fixed follow-up intervals. An example is a study of the association between stressful life events and the onset of depression. Stress exposure is assessed through the frequency and intensity of stressful life events occurring each month. Both the number of events and the intensity of each event at each measurement occasion are informative about the underlying severity of stress over time. One might hypothesize that people that approach the onset of a depressive episode have worse stress profiles than the controls, reflected by both more frequent and more intense stressors. We propose models to analyze data collected repeatedly on both the frequency of an event and its severity when both of these are informative about the underlying latent severity. Maximum likelihood estimators are developed, and simulations with small to moderate sample sizes show that the estimators also have good finite sample properties, and they are robust against misspecification of the model. This method is applied to a psychiatric data set.

Keywords

clustered data; repeated measures; informative cluster size; joint modeling; recurring episodes

1. Introduction

Cluster correlated data are often collected in medical research. For example, in some clinical trials, the response to treatment may be assessed repeatedly over time and an individual's observations are likely to be dependent; similarly, in a dental study, the teeth from the same person are clustered together. In recent years, there has been an increasing interest in modeling clustered data and a number of statistical techniques that account for the

[*]Correspondence to: Ana-Maria Iosif, Division of Biostatistics, Department of Public Health Sciences, University of California Davis, Davis, CA 95616, U.S.A.
[†]aiosif@ucdavis.edu

dependence between the members of the same cluster have been developed. Generalized estimating equations (GEE), proposed in the 1986 landmark papers by Liang and Zeger [1, 2], are among the most widely used methods of this type.

Traditional techniques for analyzing clustered data assume that the sizes of the clusters are not statistically related to the outcome of interest. However, in some contexts, this assumption may not be true, and the cluster size may be informative. For example, in the dental study described in Hoffman *et al*. [3], if the outcome of interest is the periodontal disease status of each tooth, more advanced disease status may be correlated with having fewer teeth. This phenomenon, where the response across the members of a cluster is related to the size of that cluster, is known as *informative* (or *nonignorable*) *cluster size* and failure to account for it may lead to biased inference. Recent literature has introduced models more appropriate for handling the possible informativeness of the cluster size. Hoffman *et al*. [3] proposed within-cluster resampling (WCR) as an approach for generalized linear models when the cluster size is informative. Their straightforward but computationally intensive strategy is to randomly sample (with replacement) one observation from each cluster to generate a subsample of independent observations and analyze it via standard univariate techniques. This procedure is repeated a large number of times to produce a series of data sets, and the WCR parameter estimates are obtained by averaging the resampled estimates. WCR yields consistent and asymptotically normal estimators, and a consistent estimator for the variance is provided. Williamson *et al*. [4] introduced an alternative to WCR using a GEE that is weighted inversely with the cluster size, and Wang *et al*. [5] showed it to be preferable to WCR and GEE for clustered longitudinal data. Further improvements were proposed when the correlation matrix is available, and the minimum cluster size is greater than one [6]. Chen *et al*. investigated robustness to misspecification of the cluster size distribution in the context of linear mixed models, when the cluster size is informative [7]. Follmann *et al*. [8] generalized WCR to broader types of clustered data when a method for analyzing independent data is available and termed their method *multiple outputation* and later introduced exact inference for these settings when within-cluster correlation is not of direct interest [9]. More recently, Cong *et al*. [10] and Williamson *et al*. [11] modeled correlated survival data when cluster sizes may be informative to the risk of the outcome.

A different way of conceptualizing clustered data with informative cluster size is to treat the measurements at the cluster level as exchangeable entries in a multivariate vector with a random length given by the number of members in the cluster. In a clinical trial of a migraine drug, the number of migraine episodes and the severity of each episode are collected over a follow-up period. Both the number of migraine episodes and their severity may be informative about the drug efficacy. In this approach, termed *random length data* by Barnhart and Sampson [12], one jointly models the size of the cluster (the number of episodes) and the potential dependent continuous outcomes (the severity of the episodes), which they assume to follow a multivariate normal distribution. One of the advantages of this technique is the ability to include clusters of size zero in the analysis. A traditional analysis, taking into account just the migraine severity, would exclude the clusters with no members and might lead to zero-length bias. However, having no migraines is highly

informative about the effectiveness of the drug. This likelihood-based method was later extended to include covariates [13] and to accommodate ordinal severity measures [14].

In this paper, we consider clustered data that are collected repeatedly (over conditions or over time) and model repeated measures clustered data when the cluster size is informative. This type of data can arise in a variety of settings, when information is gathered repeatedly on both the frequency of an event and its severity. In our experience, when data are collected longitudinally, the exact time of events may not have been obtained, and the information about the events is collected at fixed follow-up intervals. A change in the underlying condition severity is reflected in simultaneous changes in both the number of events and the severity of each event. Because both the frequency and the severity are important, in order to appropriately determine the treatment effect, one needs to jointly model the number of events and their associated severity measures. An example of where this model is effective is a clinical trial of a migraine drug where the data are recorded monthly. In addition to the total number of migraine episodes occurring during the respective month, the pain levels corresponding to each migraine are reported as well. Both the number of migraine episodes and the pain level of each migraine at each measurement occasion are informative about the treatment effect over time. If the drug is efficacious, the patients who received the active treatment are expected to have better pain profiles than the placebo patients; in time, they will have fewer and less severe migraines as compared with placebo.

We refer to this type of data, when individuals are observed repeatedly and their multivariate random length measurements are recorded as a series of observations as *repeated clustered data with informative cluster size*. We use the term *repeated* to indicate that each individual is measured repeatedly (under different conditions, at different assessment times, etc.). The term *clustered data* implies that the outcome for an individual recorded at a measurement occasion is in fact a cluster of severities, and the size of this cluster is a random variable determined by the number of events experienced during that follow-up interval. Finally, we use the term *informative cluster size* to point out the relatedness of the number of events to the severity measures experienced within a measurement interval.

Models for *repeated clustered data* are necessarily complex because they must consider three types of dependence within a subject: first, between continuous severity measures at a single measurement occasion; second, between severities at different measurement occasions; and third, between the number of events experienced at different measurement occasions. We provide a general likelihood-based framework for modeling repeated measures clustered data with informative cluster size.

Our research was motivated in part by the life events and difficulties schedule (LEDS) data set, collected as part of a larger study at the University of Pittsburgh Western Psychiatric Institute and Clinic. This data set contains information about stressful life events experienced by adolescents prior to a depressive episode, assessed using structured instruments [15]. In addition to the number of life events experienced over a 1-year period before the onset of the depressive episode, the degree of severity of each event was recorded. A detailed description of this study is provided in Section 2.

This remainder of the paper is organized as follows. In Sections 2 and 3, respectively, we detail our motivating problem and introduce our notation and formally describe the model. In Section 3.3, we develop an estimation strategy and derive the asymptotic properties for the proposed estimators. Simulations examine the finite sample behavior of the estimators in Section 4 and their robustness in Section 5. Our method is applied to the motivating psychiatric data in Section 6, and a brief discussion is in Section 7. Proofs of the asymptotic properties of our estimators are presented in the Appendix.

2. Motivating data

The data that motivated this research were collected as part of a larger study at the University of Pittsburgh's Western Psychiatric Institute and Clinic. These data contain information about adverse life experiences in adolescents prior to a major depressive episode. The objective of the larger study was to examine life events and long-term difficulties occurring in adolescents prior to and during a recent depressive episode. Stress exposure was assessed using the investigator-administrated LEDS [16], adapted for use with adolescents [17]. Life events are acute in nature, typically unfolding over a short period (e.g., fight with a boyfriend and death of a pet), whereas difficulties are chronic, being present for at least 4 weeks (e.g., a tumultuous relationship, with frequent fighting and arguing or serious ongoing problems with a parent). The severity of each stressor was rated on a 4-point scale with higher values reflecting more stress (4-marked, 3-moderate, 2-some, 1-little, or none). In this paper, we focus only on the acute life events in the LEDS data, treating their severity levels as continuous measurements because the literature on life stress typically uses normal distribution to analyze ordinal life events measures, usually via regression models [18,19].

Previous literature [20] has suggested a causal relationship between stressful life events and the onset of a major depressive episode. Major depressive disorder (MDD) is a serious psychiatric condition that afflicts millions of people each year and the leading cause of disability worldwide, according to the World Health Organization [21]. The prevalence of MDD in US youth (ages 12 to 17 years) is about 5% [22]. Given the personal, social, and economic costs associated with this condition, there is a tremendous clinical interest in understanding the role that stressful life events may play in the onset of depression. Traditional life events analyses would run two separate analyses for the number of events and their severities. We hypothesize that the underlying stress ‘severity’ over time drives both the number of events reported and also their severity; and this group severity differs between people with MDD and controls. Thus, we develop methodology that models jointly the two outcomes longitudinally and obviates the need for separation of information. The data set we consider contains 32 women between the ages of 13 and 18 years who have become diagnosed with MDD and 30 group matched (on age and ethnicity) normal control (NC) women. We examine the occurrence of the acute life events collected over four 3-month intervals (quarters) in the year prior to the onset of the depressive episode in adolescents with MDD episodes and during a comparable ‘linked’ period in NCs. The expectation is that in comparison with NCs, adolescents approaching the onset of an MDD episode would have reported progressively more acute life events in the year prior to the diagnosis and that these life events would be more severe.

3. Model description and estimation

In this section, we develop a methodology to model and analyze *repeated clustered data* where there are multiple experimental groups. Assume that group i is characterized by the parameters $\mu_{i1},\ldots, \mu_{iT}$, reflecting the underlying condition severity at measurement occasions 1, 2,…, T. Individuals are observed repeatedly, and at each measurement occasion, the data for a person consist of the number of events he/she experienced and the cluster of corresponding event severities. The underlying condition severity drives both the number of events and their severities, and thus, the distributions of the cluster sizes and multivariate severities share parameters. The proposed model defines parametric structures for the joint distribution of the cluster sizes and event severities for individuals in multiple groups (e.g., treatments) observed at repeated intervals and for within-individual covariance structure. The shared parametrization exploits this notion, improving efficiency over separate parameterizations. We use a log-linear functional dependence between the two mean structures and allow all groups to share scaling parameters, whereas μ's, reflecting the underlying condition severity, are group specific. The model is motivated by our belief that these scaling parameters are parameters of the process linking the underlying condition to the number of events. Treatment (or group) does not affect the relationship between number and intensity of events, just the severity of the underlying condition producing events, and therefore, the scaling parameters should remain the same, regardless of group. We then obtain maximum likelihood estimates for all parameters and derive their asymptotic properties.

3.1. Notation

This model is appropriate for studies involving multiple treatments or groups. Each individual j, $j = 1,\ldots,n_i$, from group i, $i = 1,\ldots,m$ is observed on T different occasions. At each measurement occasion $t = 1,\ldots, T$, individual j from group i reports a random number of events K_{ijt}, and the severities of these K_{ijt} events can be recorded as a vector with exchangeable entries $\boldsymbol{X}_{ijt}$. Thus, all the data for individual i can be condensed into a $\sum_{t=1}^{T} K_{ijt}$-dimensional vector $\boldsymbol{X}_{ij}$, $\boldsymbol{X}'_{ij} = (\boldsymbol{X}'_{ij1},\ldots, \boldsymbol{X}'_{ijT})$ and the corresponding T-dimensional vector of random numbers of events $\boldsymbol{K}_{ij} = (K_{ij1},\ldots, K_{ijT})$, with $i = 1,\ldots, m$, $j = 1,\ldots, n_i$. Let $\boldsymbol{k}_{ij} = (k_{ij1}, \ldots, k_{ijT})$ be a realization of the T-dimensional vector $\boldsymbol{K}_{ij}$.

We first specify the model for the cluster sizes $\boldsymbol{K}_{ij}$, choosing a Poisson distribution as a natural model for count data. We assume, for simplicity, that the mean cluster size is fully determined by the severity of the underlying condition at time t for treatment i, μ_{it}, and the scaling parameters δ and γ. Specifically, we assume that for each group $i = 1,2,\ldots, m$, and for each individual j, $j = 1,\ldots, n_i$, the cluster sizes $K_{ij1}, K_{ij2}, \ldots, K_{ijT}$ are independent and distributed Poisson (λ_{it}), where $\lambda_{it} = \exp(\delta + \gamma\mu_{it})$, $t = 1,\ldots,T$. This log-linear model provides a flexible and meaningful way to model the relationship between the number of events and their severities. The parameters μ_{it} directly characterize the mean severity in group i at measurement occasion t. γ acts as a scaling parameter and also controls the association of μ's with the cluster sizes. Because the Poisson (λ_{it}) distribution is stochastically increasing in λ_{it}, the previous model describes the following clinical

phenomena: $\gamma > 0$ implies both more severe and more frequent events. Parameter $\gamma < 0$ implies that the larger the underlying parameter μ is, the smaller the expected number of events (fewer but worse events). If $\gamma = 0$, then underlying condition severities μ_{it} have no effect on the distribution of the cluster sizes.

We then specify a multivariate normal distribution for the intensity of symptoms at each event. We use a patterned exchangeable correlation structure to describe the association of severities within-person, allowing severities recorded at the same measurement occasion to be more highly correlated than severities recorded at two different measurement occasions. This within-person dependence in the severities is modeled using the same parameters σ^2, ρ, and ρ^*, regardless of group. Clusters with $k_{ijt} = 0$ will contribute no events to the vector of observed intensities. They will, however, contribute information about treatment efficacy through the cluster sizes. To keep the notation simple, we write the model for nonzero components, but we keep in mind that if some of these components are zero, we have to account for this. The random vector of severities $\boldsymbol{X}_{ij}$, with cluster sizes $\boldsymbol{K}_{ij} = (K_{ij1}, \ldots, K_{ijT})$ for the j-th individual in the i-th group, has the conditional distribution

$$\boldsymbol{X}_{ij} | (\boldsymbol{K}_{ij} = (k_{ij1}, \ldots, k_{ijT})) \sim \mathbf{MVN}_{\sum_{t=1}^{T} k_{ijt}} \left(\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*) \right), \quad (1)$$

with mean

$$\boldsymbol{\mu}_{\boldsymbol{k}_{ij}} = \boldsymbol{\mu}_{k_{ij1}, \ldots, k_{ijT}} = \left(\mu_{i1} \boldsymbol{e}'_{k_{ij1}} \mu_{i2} \boldsymbol{e}'_{k_{ij2}} \ldots \mu_{iT} \boldsymbol{e}'_{k_{ijT}} \right) \quad (2)$$

and covariance

$$\sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*) = \sigma^2 \begin{pmatrix} \mathbf{R}_{k_{ij1}}(\rho) & \rho^* \mathbf{J}_{k_{ij1}, k_{ij2}} & \cdots & \rho^* \mathbf{J}_{k_{ij1}, k_{ijT}} \\ \rho^* \mathbf{J}_{k_{ij2}, k_{ij1}} & \mathbf{R}_{k_{ij2}}(\rho) & \cdots & \\ \cdots & \cdots & \cdots & \\ \rho^* \mathbf{J}_{k_{ijT}, k_{ij1}} & \rho^* \mathbf{J}_{k_{ijT}, k_{ij2}} & \cdots & \mathbf{R}_{k_{ijT}}(\rho) \end{pmatrix}, \quad (3)$$

where $\boldsymbol{e}_k$ denotes the k dimensional vector with all entries equal to 1, $\mathbf{J}_{k_1, k_2}$ denotes the $k_1 \times k_2$ dimensional matrix with all entries equal to 1, and $\mathbf{R}_k(\rho)$ is the intraclass correlation matrix of dimension k. To ensure that the matrix $\mathbf{S}_{k_1, \ldots, k_T}(\rho, \rho^*)$ is positive definite for all possible values $(k_1, \ldots, k_T)$, we assume that $0 \leq \rho^* \leq \rho < 1$.

For entries corresponding to nonzero k_{ijt}, the mean is given by μ_{it}. For instance, an individual from group i who reports two events at each measurement occasion has a $2T$-dimensional mean severity vector $(\mu_{i1}, \mu_{i1}, \mu_{i2}, \mu_{i2}, \ldots, \mu_{iT}, \mu_{iT})$. But if he reports no events at time 2, but two events at all the other measurement occasions, the mean is given by the $(2T - 2)$-dimensional vector $(\mu_{i1}, \mu_{i1}, \mu_{i3}, \ldots, \mu_{i3}, \ldots, \mu_{iT}, \mu_{iT})$. The covariance matrix $\mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*)$ has diagonal elements $\mathbf{R}(\rho)$ for entries corresponding to the nonzero components of $\boldsymbol{k}_{ij}$

and off diagonal components corresponding to entries with both k_{ijt} and $k_{ijt'}$ nonzero equal to ρ^*. For example, an individual who reports two events at each of the first two measurements and zero events at the rest of the measurements would have a 4 × 4 covariance matrix equal to

$$\sigma^2 \begin{pmatrix} \mathbf{R}_2(\rho) & \rho^* \mathbf{J}_{2,2} \\ \rho^* \mathbf{J}_{2,2} & \mathbf{R}_2(\rho) \end{pmatrix}.$$

We finally assume that data from different individuals are independent. Thus, $(\boldsymbol{K}_{i_1 j_1}, \boldsymbol{X}_{i_1 j_1})$ and $(\boldsymbol{K}_{i_2 j_2}, \boldsymbol{X}_{i_2 j_2})$ for individual j_1 from group i_1 and j_2 from group i_2, respectively, are independent for $(i_1, j_1) \neq (i_2, j_2)$.

3.2. Model features

Denoting the $(mT + 5)$ parameters collectively by $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, \ldots, \mu_{1T}, \ldots, \mu_{m1}, \ldots, \mu_{mT}, \sigma^2, \rho, \rho^*)'$, the parameter space for the previous model is $\boldsymbol{\Theta} = \{\boldsymbol{\theta} | -\infty < \delta, \gamma, \mu_{11}, \ldots, \mu_{mT} < \infty, \sigma^2 > 0, 0 \leq \rho^* \leq \rho < 1\}$.

The support of the cluster sizes includes zero. Any of the components of the vector $\boldsymbol{K}_{ij}$ may be zero. Note that, to account for observation times with $K_{ijt} = 0$ (no observed events), the vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{S}$ will have corresponding structural empty cells (no corresponding entry for that time point in both the mean vector and the covariance matrix). Thus, the actual working dimension of $\boldsymbol{\mu}$ and $\mathbf{S}$ will reflect only the nonzero clusters, but the linkage to K_{ijt} by time of observation is preserved. Consider that the individuals described previously reports the same number of events, but this time, he has 0 events recorded at the first time measurement and two events recorded at each of the next two time points. Conditional on the number of events experienced during all measurement occasions, the covariance for the multivariate severities looks the same, being equal to

$$\sigma^2 \begin{pmatrix} \mathbf{R}_2(\rho) & \rho^* \mathbf{J}_{2,2} \\ \rho^* \mathbf{J}_{2,2} & \mathbf{R}_2(\rho) \end{pmatrix},$$

but one needs to take into account the fact that this structure corresponds to the first two time measurements for the first case and to the measurements occasions 2 and 3 in the second case.

This is one of the main difficulties in handling the model, because not only the number of events but also the number of blocks that constitute the mean and covariance structures in (2) and (3) can change from person to person. Thus, strict attentiveness and a significant amount of bookkeeping need to be conveyed in working with the conditional density functions for the multivariate severity measurements.

3.3. Likelihood framework and estimation

We can write the joint density for the j -th individual's data in the i -th group as $f(\boldsymbol{x}_{ij}, \boldsymbol{k}_{ij}) = \mathbf{P}(K_{ij1} = k_{ij1},\ldots,K_{ijT} = k_{ijT}) f(\boldsymbol{x}_{ij} \mid k_{ij1},\ldots, k_{ijT})$, provided that $k_{ij\cdot}=\sum_{t=1}^{T} k_{ijt} \geq 1$. We denote by $f(\cdot|k_{ij1},\ldots,k_{ijT}) = f(\cdot|\boldsymbol{k}_{ij})$ the density of the $k_{ij\cdot}$-dimensional multivariate normal variable from $\mathbf{MVN}_{k_{ij\cdot}}(\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}, \sigma^2\mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*))$. If a person does not have any events at any of the measurement occasions, the likelihood reduces to the first term in the previous expression.

Then the log-likelihood of the entire data can be written as

$$l(\theta)=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\mathbf{p}(K_{ij1}=k_{ij1},\ldots,K_{ijT}=k_{ijT})+\sum_{i,j}\sum_{k_{ij\cdot}\geq 1}\log f(\boldsymbol{x}_{ij}|k_{ij1},\ldots,k_{ijT}),$$

where the parameter vector is $\theta = (\delta, \gamma, \mu_{11},\ldots, \mu_{mT}, \sigma^2, \rho, \rho^*)'$.

Estimation will be carried out by maximum likelihood, under the assumption that the appropriate number and distribution of events are observed. For instance, if no events, across all people and all measurement periods are observed (i.e., no $k_{ijt} > 0$), then none of the parameters are estimable. If all the people in the sample experience at most one event, then the parameter ρ is not estimable. If all the people in the sample experience events only at the same measurement occasion, then the parameter ρ^* is not estimable. For large sample sizes, these difficult cases become increasingly rare. As expected, there is no closed form solution for the maximum likelihood estimator (MLE) $\hat{\theta_n}$ for θ, and its exact distribution is not available. We derive the asymptotic distribution for $\hat{\theta_n}$ in Section 3.4. Numerical methods for maximization are discussed in Section 4.2.

3.4. Asymptotic distribution of the maximum likelihood estimators

Using a generalization of the principle of adding the information in [12], the information $\boldsymbol{I}(\theta)$ about the parameter θ contained in a single observation $\boldsymbol{X_{ij}}$ with cluster sizes $\boldsymbol{K}_{ij} = (K_{ij1}, \ldots, K_{ijT})$ from group i can be written as the sum of the information about θ contained in the cluster sizes and the information about θ contributed by the vectors of severities, over all possible cluster sizes (under the assumption that at least one event is observed, i.e., $k_{\cdot}=\sum_{t=1}^{T} k_t \geq 1$).

$$\boldsymbol{I}^{(i)}(\boldsymbol{\theta})=\boldsymbol{I}_i^*(\boldsymbol{\theta})+\sum_{k_{\cdot}\geq 1}\mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K}_{ij}=\boldsymbol{k})\boldsymbol{I}_i(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}_i^*(\boldsymbol{\theta})$ is the information matrix about θ contained in the cluster sizes $\boldsymbol{K}_{ij} = (K_{ij1},\ldots, K_{ijT})$ and $\boldsymbol{I}_i(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}_{ij}|\boldsymbol{K}_{ij} = \boldsymbol{k} = (k_1,\ldots, k_T)$.

Let us denote by $\boldsymbol{I}_n(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the $n=\sum_{i=1}^{m} n_i$ independent observations from the multiple group model, $\boldsymbol{X}_{ij}$ with cluster sizes $\boldsymbol{K}_{ij} = (K_{ij1},\ldots, K_{ijT})$, $i = 1, \ldots,m$, $j = 1,\ldots,n_i$.

Then the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ contained in these n independent observations from the multiple group model is

$$\boldsymbol{I}_n(\boldsymbol{\theta})=\sum_{i=1}^{m} n_i \boldsymbol{I}_i^*(\boldsymbol{\theta})+\sum_{i=1}^{m} n_i \sum_{k_\cdot \geq 1} \left[\prod_{t=1}^{T} e^{-e^{\delta+\gamma\mu_{it}}} \frac{e^{k_t(\delta+\gamma\mu_{it})}}{k_t!}\right] \boldsymbol{I}_{\boldsymbol{i}}(\boldsymbol{\theta}|\boldsymbol{k}).$$

We show in the Appendix that $\boldsymbol{I}_n(\boldsymbol{\theta})$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta})=\begin{pmatrix} \boldsymbol{I}_n(\delta,\gamma,\mu_{11},\ldots,\mu_{mT}) & \mathbf{O}_{(mT+2)\times 3} \\ \mathbf{O}_{3\times(mT+2)} & \sum_{i=1}^{m} n_i \sum_{k_\cdot \geq 1} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K}_{ij}=\boldsymbol{k})\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*) \end{pmatrix}.$$

where $\mathbf{O}_{n\times m}$ denotes the $n \times m$-dimensional matrix of zeros and the components on the main diagonal are defined by (A14) and (A3) – (A8) in the Appendix, respectively.

Applying a general result on the efficiency of MLEs for random length data (Theorem A.3.2 in [23]), we can derive the asymptotic distribution for $\hat{\boldsymbol{\theta}}_n$, the MLE. The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}_n$ is obtained as the inverse of the previous information matrix and is estimated by $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$.

Let $\hat{\boldsymbol{\theta}}_n=(\hat{\delta}_n,\hat{\gamma}_n,\hat{\mu}_1^{(n)},\ldots,\hat{\mu}_{11}^{(n)},\ldots,\hat{\mu}_{1T}^{(n)},\ldots,\hat{\mu}_{mT}^{(n)},\hat{\sigma}_n^2,\hat{\rho}_n,\hat{\rho}_n^*)'$ be the MLEs for a sample of size n from the multiple group model. If $n_i/n \to \eta_i$ with $0 < \eta_i < 1$ as $n \to \infty$, then

1. $\hat{\boldsymbol{\theta}}_n$ is consistent.
2. $$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{L} \mathbf{MNV}_{mT+5}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

   where

$$\boldsymbol{I}(\boldsymbol{\theta})=\begin{pmatrix} \boldsymbol{I}(\delta,\gamma,\mu_{11},\ldots,\mu_{mT}) & \mathbf{O}_{(mT+2)\times 3} \\ \mathbf{O}_{3\times(mT+2)} & \sum_{i=1}^{m} \eta_i \sum_{k_\cdot \geq 1} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K}_{ij}=\boldsymbol{k})\boldsymbol{I}_k(\sigma^2,\rho,\rho^*) \end{pmatrix},$$

$$\boldsymbol{I}(\delta,\gamma,\mu_{11},\ldots,\mu_{mT})$$
$$=\sum_{i=1}^{m}\eta_i\begin{pmatrix}\mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times(i-1)T} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times(m-i)T}\\ \mathbf{O}_{(i-1)T\times 2} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{O}_{(i-1)T\times T} & \mathbf{O}_{(i-1)T\times(m-i)T}\\ \mathbf{G}_{12}^{(i)'} & \mathbf{O}_{T\times(i-1)T} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times(m-i)T}\\ \mathbf{O}_{(m-i)T\times 2} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{O}_{(m-i)T\times T} & \mathbf{O}_{(m-i)T\times(m-i)T}\end{pmatrix}$$
$$+\frac{1}{\sigma^2}Diag\left(\mathbf{O}_{2\times 2},\eta_1\sum_{k.\geq 1}\prod_{t=1}^{T}e^{-\lambda_{1t}}\frac{{\lambda_{1t}}^{k_t}}{k_t!}\sum_{\boldsymbol{k}}^{-1},\ldots,\eta_m\sum_{k.\geq 1}\prod_{t=1}^{T}e^{-\lambda_{mt}}\frac{{\lambda_{mt}}^{k_t}}{k_t!}\sum_{\boldsymbol{k}}^{-1}\right),$$

the elements of the matrix $\boldsymbol{I}_{\mathbf{k}}$ (σ^2, ρ, ρ^*) are defined earlier, and the elements in $\boldsymbol{I}(\delta, \gamma, \mu_{11},\ldots, \mu_{mT})$ are given in the Appendix by (A10) and (A2).

4. Simulation study

In this section, we report the results of a simulation study conducted to evaluate the finite sample properties of our estimators in realistic situations with sample sizes in a small to medium range. Motivated by our LEDS data set, we simulated data that closely resemble its structure. Several different scenarios and sample size configurations are analyzed, with data generated according to the multiple group model described in Section 3.

We evaluate the performance of the proposed estimators by examining three properties: accuracy, precision, and ability to make valid inference. Varying the sample sizes allows us to tackle the question of how many subjects are necessary in order for the large-sample theory to produce the desired results.

4.1. General framework and quantities computed

Four different scenarios were created to explore various parameter configurations, representing a 2×2 factorial design. The first factor involves the relationship between the severities and the number of events. We believe the two interesting levels for this factor are those reflecting contradictory information from the number of events and severities. In the first level, μ_{it}, the expected severities are similar for the two groups, but the differences between mean number of events are exaggerated by a large γ and small δ In the second level, the μ's are set to be spread out across groups, but small γ and large δ yield very similar mean numbers of events. The second factor refers to the pattern of group differences over time. The first level has parallel time trajectories for severity in the two groups, whereas in the second level, mean severity trajectories have different slopes: one increasing and one decreasing over time.

As in the LEDS setup, we mimic a study with two different treatment groups and the same number of subjects per treatment group. All subjects are followed four measurement periods. We keep the same values for σ^2, ρ, and ρ^* across all scenarios, and we specify the remaining parameters reflecting the four different circumstances over a range of plausible values. The complete choices of parameters for simulations are given in Table I. For each simulated scenario, 1000 data sets were generated, with n = 20, 50, and 100 subjects per group.

The number of events k_{ijt} experienced by individual $j = 1,\ldots, n$ in group $i = 1, 2$ at measurement occasion $t = 1,\ldots, 4$ was generated using independent Poisson (λ_{it}) probabilities, with $\lambda_{it} = e^{\delta+\gamma\mu_{it}}$. If all the components of the vector of cluster sizes $\boldsymbol{k}_{ij} = (k_{ij1}, k_{ij2},\ldots, k_{ij4})$ are zero, we have an empty cluster of severities. Otherwise, we generate the severities $\boldsymbol{X}_{ij}$ using a multivariate normal distribution with mean and variance given by (2) and (3), respectively. We analyze each data set using our multiple group model and compute the ML estimator of the parameter $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11},\ldots, \mu_{14}, \mu_{21},\ldots, \mu_{24}, \sigma^2, \rho, \rho^*)$ and its theoretical asymptotic variance as described in Section 3.3. We report the average of the bias, the standard error (SE), and the coverage probabilities of the 95% confidence intervals for each parameter setup. Specifically, we evaluate the accuracy of the proposed estimators by computing their empirical bias and the precision by reporting their SE. The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ is obtained as the inverse of the information matrix and is estimated by its plug-in estimator, $\boldsymbol{I}_n^{-1}(\hat{\theta})$. Using the estimated variance, we construct 95% confidence intervals for all the parameters. To evaluate the ability to make valid inference using the proposed estimators, we report the percentage of 95% confidence intervals that contained the true value of the parameter.

4.2. Considerations for numerical maximization

Using a shared parameter model in both the specification of the distribution of the cluster sizes and the conditional distribution of the vectors of severities given the cluster sizes increases the efficiency, by pooling information from both outcomes. The cost is an increase in the computational burden, because the shared parameters need to be estimated simultaneously from the two models.

Specifically, there are several difficulties that arise in modeling repeated clustered data with informative size. First, the full likelihood approach is computationally complex when the number of events is large, because higher dimension matrices are involved in the conditional distribution of the multivariate severity measures. Second, the complexity increases as the number of repeated measurements becomes larger and the number of parameters to be estimated increases. Because the number of time points with quantitative measures changes with every person, the mean and covariance structures for the distribution of the severities change. Moreover, numerical computation of the information matrix involves summation over all possible values of the cluster sizes. For T measurement times that means summation of matrices over all the possible values of a T-dimensional vector of cluster sizes. If the average number of events is large, in numerically computing the information matrix, one needs to set the maximum values that the cluster size can have quite high to ensure that the corresponding probabilities in the right tails of the Poisson distributions are zero.

Finally, a third difficulty relates to modeling of slopes. If all the μ's are equal, the parameters δ and γ are not identifiable. When data are generated from groups with means not well separated, the likelihood can be flat over certain regions and the function used to numerically maximize the log-likelihood converges to strange solutions for the parameters δ and γ. However, even in these cases, the average severity and number of events are estimated correctly (Section 4.3).

All the simulations were implemented in R statistical programming language [24], using its function *nlm* to carry out the maximization of the log-likelihood. This function minimizes the negative log-likelihood using a Newton-type algorithm and requires specification of initial values for the parameters. We obtain good initial estimates for the parameters μ_{it} reflecting the underlying disease status in group i at measurement occasion t using $\mu_{it}^{(0)}=\sum_{j=1}^{n}\overline{x}_{ijt}\boldsymbol{I}(k_{ijt}>0)/\sum_{j=1}^{n}\boldsymbol{I}(k_{ijt}>0)$. We then find the starting values for δ and γ by fitting a Poisson regression model, $\log(\lambda_{ijt})=\delta+\gamma\mu_{it}^{(0)}$ $i = 1, 2$ $j = 1, 2,\ldots, n$ $t = 1,\ldots, 4$, where $\lambda_{ijt} = E(k_{ijt})$. We solve for σ^2 in the corresponding likelihood equation to get $\sigma^2(\rho, \rho^*)$ and substitute this value in the formula for the conditional likelihood to get an expression that depends only on the unknown ρ and ρ^* Maximizing this likelihood with respect to ρ and ρ^* by using *nlm* with 0 as starting values for both parameters yields the initial values $\rho^{(0)}$ and $\rho^{*(0)}$. Plugging these values into the expression of $\sigma^2(\rho, \rho^*)$ gives the initial value $\sigma^{2(0)}$. If any of the initial values for σ^2, ρ, or ρ^* are negative, they are assigned the value zero. With these initial estimates, the optimization proceeds using *nlm*.

4.3. Simulation results

Results for the simulations across the four different scenarios and three sample sizes choices are summarized in Table II (bias), Table III (SE), and Table IV (coverage of the 95% confidence intervals). The main parameters of interest are $\mu_{11},\ldots, \mu_{24}$, reflecting the underlying condition severity in the two groups over time. Also of interest are the parameters that model the number of the events $\lambda_{11},\ldots, \lambda_{24}$. The results of the simulations indicate that the estimates of the true μ's and λ's are unbiased even for small sample sizes, under all four different scenarios. The variance components σ^2, ρ, and ρ^* are also nearly unbiased. For small sample size, the bias of the parameter closer to the boundary (ρ^*) is slightly higher. As shown by Table II, some of the estimates for δ and γ have large bias for small sample sizes. For $n = 20$, the estimates for δ and γ are biased in both scenarios 1 and 2, but the bias tends to wash out even with a sample size as small as 50. In scenarios 3 and 4, the estimation for small sample sizes works as well as in the cases with n large. The main source of this difference in results across scenarios is the disproportion between the average severities and the variance σ^2. In the first two scenarios, the theoretical severities are not well separated across time and treatment group with respect to σ^2, as they are in the latter scenarios. In addition, in scenario 2, the severities for the two groups intersect across time, making estimation more difficult. It is important to note that even in cases where the scaling parameters δ and γ are not closely estimated, estimation for the key parameters, $\mu_{11},\ldots, \mu_{24}$ and $\lambda_{11},\ldots, \lambda_{24}$, have practically insignificant bias, even for sample sizes as small as 20.

We evaluate the efficiency of the MLEs in our proposed method and contrast it with WCR, a commonly used method for clustered data. To obtain the WCR estimates, we used 5000 resamples and a linear model with group, time, and their interaction. We present in Table III the empirical standard errors for the parameters that are meaningful in both methods (the severities) and the relative efficiency (defined as the ratio of the calculated variance of $\hat{\mu}$ under our model relative to WRC). As expected, the relative efficiency is always greater than 1, and the gain in efficiency is substantial in the scenarios in which the cluster sizes

provide more information (1 and 2). For all three sample sizes under consideration, the efficiency gains are major in both scenario 1 (range 36% to 222% across all sample sizes) and scenario 2 (range 131% to 305% across all sample sizes) and more modest in scenario 3 (range 2% to 8% across all sample sizes) and scenario 4 (range 3% to 16% across all sample sizes).

Table IV shows the empirical coverage of the 95% confidence intervals based on the large sample variance of the proposed estimators. Paralleling the results for the bias, our method yields satisfactory coverage for all the key parameters of interest, the mean severities and numbers of events, even for choices of n as small as 20 (with the smallest coverage 92%). For the scaling parameters δ and γ, the coverages are below 95% when the sample sizes are small, due to the bias. Unsurprisingly, the normal approximation is not as good for the variance components, with our small to medium sample sizes and the coverage of their confidence intervals, although reasonable, is not improving as fast with the sample size. Much larger samples than those we consider here are needed before the asymptotics apply for these parameters. However, the estimation of the primary parameters is robust and does not seem to be affected by the secondary parameters.

5. Robustness of the model

Our methodology involves explicit modeling of both the severities and the cluster sizes. In this section, we examine, through simulations, the robustness of our model. We consider three different types of model misspecification and for each of these, we simulate 1000 data sets under each scenario and sample size combination used in the simulation study described in Section 4. We report in Tables V–VIII the bias and coverage of the 95% confidence intervals of the MLEs computed using the model-based SEs described in Section 3 for the parameters $\mu_{11},\ldots, \mu_{14}$ reflecting the underlying condition severity (representing the primary interest to investigators), parameters $\lambda_{11},\ldots, \lambda_{14}$, and the scaling parameters δ and γ, for n = 50. The Supporting information[‡] contains complete information for all three sample sizes under consideration.

In the first type of distributional form misspecification, we consider that the true covariance for the severities is of a more complicated and realistic structure while assuming that the distribution of the cluster sizes was correctly specified. To explore how violations of each of the between-measurement and within-measurement occasion exchangeability assumptions affect estimation, we include two models for the covariance structure. Under the first model, we simulate data assuming that severities collected during the same observation period are correlated with $\rho = 0.5$ and allow severities from different measurement occasions to be correlated using an AR (ρ^*) structure, with $\rho^* = 0.2$. Under the second model, we maintain the exchangeability structure across measurement occasions ($\rho^* = 0.2$) but allow severities from the same measurement occasions to be correlated using an autoregressive (AR) (0.5)+ ρ^* structure.

[‡]Supporting information may be found in the online version of this article.

The second type of model misspecification focuses on the assumption of independence of the cluster sizes across measurement occasions while assuming that the distribution of the severities was correctly specified. We simulate cluster sizes reflecting the number of events experienced by a person across measurement occasions from a multivariate Poisson distribution and allow the cluster sizes to be correlated across measurement occasions. We use a simplified model with just one covariance term for all pairs of cluster sizes [25]. Marginally, each of the cluster sizes follows a Poisson distribution with parameter (λ_{it}), and the covariance between all pairs of random cluster sizes across measurement occasions is λ_0. Table VII presents the results of the simulations for $\lambda_0 = 1$ and $n = 50$.

The third misspecification incorporates extra-Poisson variation in the distribution of cluster sizes. Specifically, we are interested in the behavior of the MLE when the true cluster size model is negative binomial (NB), whereas the working model is Poisson (λ_{it}). The mean and variance of NB (α, λ_{it}) are λ_{it} and $\lambda_{it}(1+\frac{\lambda_{it}}{\alpha})$, respectively. Thus, α controls how overdispersed the NB distribution is compared with α Poisson distribution with mean λ_{it}. Table VIII presents the results of the simulations for $\alpha = 10$ and $n = 50$.

The results displayed in Tables V–VIII show that when our proposed model is applied to data generated from these misspecified models, there are very small biases in the estimates of the mean severities for $n = 50$. The relative bias (absolute bias divided by the true value) for these parameters was less than 1%. Similarly, there was very little loss of accuracy in estimating the mean number of events (the relative bias for λ's was smaller than 2%). This pattern was consistent across sample sizes (Tables A1–A4 in the Supporting information). Parameters δ and γ have larger bias and lower coverage in scenarios 1 and 2, as in the original simulation study.

When the covariance structures (of either the severities or cluster sizes) was misspecified (Tables V–VII, A1–A3), the coverage probabilities of the key parameters (μ and λ) are close to 95%. As shown in Table VIII, the extra-Poisson variation considered does not seem to impact the coverage of the mean severities, but as expected (because the asymptotic variance estimates used to construct the confidence intervals for the cluster sizes are based on the Poisson assumption), the coverage for the λ's is not as good, and it can be as low as 80%.

For the three types of violation of the assumption considered, we found that there is little loss of accuracy in estimating the mean severities and cluster sizes, even for small sample sizes. Both misspecification of the cluster size distributions and ignoring the within-person dependence for the severities or cluster sizes yield estimates with good coverage probabilities for the severities. When substantial extra-Poisson variation is incorporated in the distribution of cluster sizes, the bias in λ's is still very small, but there is undercoverage of the confidence intervals. For investigators whose primary interest is in estimating the main severities, the results of the simulation suggest that even though our model makes structural assumptions, the three types of departure from the model assumption considered in this section may not have a major influence in the estimation. If the main focus is on estimating the number of events, then misspecifications of the covariance structures will also

have little impact on estimation, but the Poisson distribution assumption needs to be verified using diagnostic tools before using this model.

6. Application to life events and difficulties schedule data

For the 32 MDD adolescents, the number of life events experienced in a quarter ranged from 70 to 92, which amounted to a range of 2.2 to 2.9 events per person. The corresponding range for the NC group was 41 to 64 (1.4 to 2.1 events per person). During the year, 20% of the clusters of severities in the MDD group and 30% in the NC group had size 0. The number of life events reported during a quarter by a single person ranged from none to 11 in the MDD group and from none to six in the NC group. As detailed in Table IX, acute life events were common in the year prior to depression onset, with 75% of the MDD and 57% of the NC adolescents reporting at least one stressful event at every assessment time. In general, a greater percent of the MDD adolescents reported at least two acute events, and this was consistent across the four assessments. The right panel of Figure 1 depicts the average severity for the events experienced by the two groups during the year preceding the onset of depression. The left panel shows the average number of events reported. More than 50% of the life events reported by the NC group at any of the assessment times had severity 1, whereas the majority of the life events reported by the MDD group had severity 2 or higher.

The two outcomes recorded for each subject are the number of acute life events and the severity of each of them. These outcomes are assessed quarterly, over a year. Each quarter, the data for an individual participant are a cluster of event severities with the size given by the number of events the participant experienced during that quarter of year. We denote by $\mu_{11},\ldots, \mu_{14}$ the parameters reflecting the underlying stress severity during the four quarters of the year before the onset of depression in the MDD group. Similarly, we denote by $\mu_{21}, \ldots, \mu_{24}$ the means in the NC group. The model introduced in Section 3 is applicable now to this data set. We have $n_1 = 32$ and $n_2 = 30$. The question we consider here is whether the underlying stress severity for the adolescents in the MDD group is escalating as they draw closer to the onset of their MDD episode.

Table X reports the solution $\hat{\theta}$ of the maximization procedure implemented in R using *nlm* and the estimated standard deviations based on $\boldsymbol{I}_n^{-1}(\hat{\theta})$, as well as the initial values for the maximization procedure.

The estimated parameter γ has a positive sign indicating a positive relationship between the average number of events and the average severity (larger severities and higher number of events). Its estimated standard deviation is large, indicating that γ is not significantly different than zero.

Another question of interest is testing if the severity profiles of the two groups are parallel. A Wald test of this composite hypothesis yields a test statistic $\chi^2 = 0.34$, which is not significant with respect to a chi-square distribution with 3 degrees of freedom. Similar results were obtained for testing whether the profiles were also coincident and equal response effects.

Using a shared parameter model allowed us to gain clinical insight into this complicated matter. We expected that both the severity and the number of events were driven by the underlying stress. Contrary to our predictions, the number of the adverse life events did not provide information about their severity. The joint distribution of the cluster sizes and severities of two groups looked similar, with the MDD group experiencing more events, but not more severe than the NC group.

7. Discussion

Many diseases may manifest with recurring episodes. The number of episodes and each episode's severity are collected repeatedly, over assessment intervals. A naive approach would be to analyze the number of episodes and their severities separately. In this paper, we propose models that allow us to analyze data gathered repeatedly on both the frequency of an event and its severity when both the frequency and the severity are important. Our shared parameter approach obviates the need for this separation of information. We propose a general ML framework and derive estimators with good asymptotic properties. Our simulations under scenarios with sample sizes in a small to moderate range show that the estimators also have good finite sample properties. We implement this method to a real life data set.

Analysis of clustered data has been a rapidly growing field in recent years, but most of the approaches do not account for the possible relationship between the outcome and the size of the cluster. Our model is appropriate in scenarios when clustered data with informative cluster size are collected repeatedly, and it has the advantage of accommodating clusters of size zero. Our initial model is relatively simple but arises naturally and permits answering more subtle questions about the relationship between frequency and severity of events. Simulation suggests that even though our model makes structural assumptions, the three types of departure from the model assumption considered may not have a major influence in the estimation of the mean severities.

For the parametric distribution of the cluster size, our choice of the Poisson distribution was based on its widespread use in modeling number of episodes and its simplicity. However, our method is not limited to this choice and can be readily generalized to other families of discrete distributions that might provide better fits, for example, the NB distribution or, more generally, a family of discrete distribution with appropriate behavior.

Building models for *repeated clustered data* is a complex task. For instance, not only does the number of events experienced by a subject change over time; the number of time points with observed quantitative measures changes with every subject. Another complicating issue is that the mean and covariance structures for the distribution of the severities change with the change of the number of time points with observed quantitative measures. When the number of repeated time measurements increases, the number of parameters that need to be estimated increases. As the number of recorded events experienced by subjects and time measurements with quantitative measure increases, the difficulty of numerically estimating the parameters in the model increases, as well. We built a platform that handles this computational challenge and can be used to generalize the models to more complex settings.

Substantial challenges remain to be addressed, providing several interesting possible directions to further this research. Our choice of the covariance structure was motivated in part by its computational convenience. In cases of longitudinal studies, an AR structure might be more appropriate. Thus, additional work is needed to accommodate data in settings where the correlation structure is more complicated than a pattern exchangeable structure. An even more general class of models would incorporate dependence between the cluster sizes across measurement occasions. In the proposed models, we treat the severity measures as continuous random variables. However, many of the severity measures encountered in practice are categorical (e.g., in LEDS data, 1='little or none', 2='some', 3='moderate', and 4='marked'), and there is interest in developing models for repeated clustered ordinal data. Finally, one can imagine that the underlying condition severity might be impacted by observed covariates, risk factors, or unobserved latent variables. Thus, incorporating covariates, including time varying or shared random effects into the models, is another potential generalization.

Although our motivation was drawn from a study examining stressful life events in adolescents, the methods we propose are useful for a broad range of chronic conditions for which repeated data are available on both the number of episodes and their severity. From a clinical standpoint, it is important to assess both these outcomes, because they both might inform about the condition severity.

These data are becoming increasingly more common, because recent health care policy changes have made widespread use of electronic health records in the USA inevitable. Electronic health records will make possible monitoring of patients with chronic conditions with episodic manifestation and tracking these episodes and their related severities over time. Methods such as ours will enable researchers to better understand differences in incidence and severity of episodes.

Acknowledgments

The authors thank the Child and Adolescent group from the Western Psychiatric Institute and Clinic at the University of Pittsburgh for permission to use the LEDS data and to Dr. Laurel Beckett for her valuable feedback on earlier drafts of this manuscript. They also gratefully acknowledge the anonymous associate editor and referee for their insightful comments and suggestions that have considerably strengthened the manuscript. This research was supported in part by Computational Resources on PittGrid (www.pittgrid.pitt.edu) and by the National Center for Advancing Translational Sciences, National Institutes of Health grant UL1 TR000002 through funds to the first author. Allan R. Sampson is supported, in part, by NIMH 084053.

Appendix A: Information matrix

Consider one observation from the multiple group model introduced in Section 3. The log-likelihood for this one observation is

$$\log f\left(\boldsymbol{x},\boldsymbol{k}\right)=\log \mathbf{P}_{\boldsymbol{\theta}}\left(K_1=k_1,\ldots,K_T=k_T\right)+I\left(k_{\cdot}\geq 1\right)\log f\left(\boldsymbol{x}|\boldsymbol{k}\right).$$

For simplicity, we dropped the indices *i* and *j* from the expression of cluster sizes and severities. Let $\boldsymbol{k}=(k_1,\ldots,k_T)$ be a realization of the *T*-dimensional vector of cluster sizes $\boldsymbol{K}$.

Some of the components of $\boldsymbol{k}$ might be zero. Provided that $\boldsymbol{k}$ has at least one nonzero component (i.e., $k_{\cdot} \geq 1$), the conditional distribution $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$, $f(\boldsymbol{x}|\boldsymbol{k})$, is a multivariate normal. For now, assume that all the components of $\boldsymbol{k}$ are nonzero. We can use a result from McCulloch *et al*. [26] who gave the expressions of the score function and information matrix for the general model under the multivariate normality assumption, $\boldsymbol{X} \sim \mathbf{MVN}(\boldsymbol{\mu}, \mathbf{V})$ with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Var(\boldsymbol{X}) = \mathbf{V}$. Under a general parametrization, each element of $\boldsymbol{\mu}$ is a function of elements of a parameter vector $\boldsymbol{\beta}$ and each element of $\mathbf{V}$ is a function of the elements of a parameter vector $\boldsymbol{\phi}$, unrelated to $\boldsymbol{\beta}$. Following the notation in [26], we have $\boldsymbol{\beta} = (\mu_{i1}, \mu_{i2},\ldots, \mu_{iT})$ and $\boldsymbol{\phi} = .(\sigma^2, \rho, \rho^*)$ where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ and $\mathbf{V} = \mathbf{V}(\boldsymbol{\phi})$. We show in [27] that the information matrix about $(\boldsymbol{\beta}, \boldsymbol{\phi}) \equiv (\mu_{i1},\ldots, \mu_{iT}, \sigma^2, \rho, \rho^*)'$ contained in $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$ is given by

$$-E\begin{pmatrix} \frac{\partial^2 l}{\partial\beta\partial\beta'} & \frac{\partial^2 l}{\partial\beta\partial\varphi'} \\ \left(\frac{\partial^2 l}{\partial\beta\partial\varphi'}\right)' & \frac{\partial^2}{\partial\varphi\partial\varphi'} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{3\times T} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \quad \text{(A1)}$$

where

$$\sum_{\boldsymbol{k}} = \text{Diag}\left(\frac{1}{\tau_{k_1}} - \rho^*, \frac{1}{\tau_{k_2}} - \rho^*, \ldots, \frac{1}{\tau_{k_T}} - \rho^*\right) + \rho^* \boldsymbol{e}_T \boldsymbol{e}_T', \quad \text{(A2)}$$

$\tau_0 = \frac{1}{1-\rho}$, and $\tau_k = \frac{k}{1+(k-1)\rho}$. The elements of the matrix in the right lower corner in (A1) are given by

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{11} = \frac{1}{2} tr\left(\frac{1}{\sigma^4}\mathbf{I}_{k_{\cdot}}\right) = \frac{k_{\cdot}}{2\sigma^4} \quad \text{(A3)}$$

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{12} = \frac{1}{2\sigma^2}\left[tr\left(\sum_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right)\right) - \tau_0 tr(\mathbf{I}_{\boldsymbol{k}_{\cdot}-T})\right] \quad \text{(A4)}$$

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{13} = \frac{1}{2\sigma^2} tr\left(\sum_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right) \quad \text{(A5)}$$

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{22} = \frac{1}{2} tr\left(\sum_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right)\sum_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right)\right) + \frac{\tau_0^2}{2} tr(\mathbf{I}_{\boldsymbol{k}_{\cdot}-T})) \quad \text{(A6)}$$

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{23} = \frac{1}{2} tr\left(\sum_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right)\sum_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right) \quad \text{(A7)}$$

$$I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{33} = \frac{1}{2} tr\left(\sum_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \boldsymbol{I}_T)\sum_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right), \quad \text{(A8)}$$

where

$$\mathbf{V}^{-1}=\frac{1}{\sigma^2}\begin{pmatrix} \sum_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{3\times T} & \tau_0\mathbf{I}_{k.-T} \end{pmatrix}.$$

We assumed all components of $\boldsymbol{k}$ to be nonzero. However, our model allows for clusters of size zero. In this instance, the corresponding entries in the matrix $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ are zero and all the previous computations have to be carried out replacing $\boldsymbol{k}$ with its subvector containing only nonzero components. The corresponding matrix $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ is actually obtained by applying the equations earlier for the subvector of nonzero cluster sizes and filling in the corresponding entries with zero so that we obtain a $T \times T$ matrix.

Let us denote $\boldsymbol{\theta}_i = (\delta, \gamma, \mu_{i1},\ldots, \mu_{iT}, \sigma^2, \rho, \rho^*)'$. Because the expression of $f(\boldsymbol{x}|\boldsymbol{k})$ involves neither δ nor γ, it follows that the information about $\boldsymbol{\theta}_i$ contributed by the vector of severities conditional on the observed cluster sizes for one individual is given by the expression

$$\boldsymbol{I}(\boldsymbol{\theta}_i|\boldsymbol{k})=\begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{T\times 2} & \frac{1}{\sigma^2}\sum_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times T} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*) \end{pmatrix}. \quad \text{(A9)}$$

Because $\log f(\boldsymbol{k})=\sum_{t=1}^{T}\left[-e^{\delta+\gamma\mu_{it}}+k_t(\delta+\gamma\mu_{it})\right]$, it is straightforward to show that $\boldsymbol{I}_i^*(\boldsymbol{\theta}_i)$, the information matrix about $\boldsymbol{\theta}_i$ contained in the random cluster sizs $\boldsymbol{K}=(K_1,\ldots, K_T)$ for one subject in group i has the expression

$$\boldsymbol{I}_{\boldsymbol{i}}^*(\boldsymbol{\theta}_i)=\begin{pmatrix} \mathbf{G}_i & \mathbf{O}_{(T+2)\times 3} \\ \mathbf{O}_{3\times(T+2)} & \mathbf{O}_{3\times 3} \end{pmatrix},$$

where the upper left-corner $(T + 2) \times (T + 2)$ dimensional matrix is given by

$$\mathbf{G}_i=\begin{pmatrix} \sum_{t=1}^{T}\lambda_{it} & \sum_{t=1}^{T}\mu_{it}\lambda_{it} & \gamma\lambda_{i1} & \ldots & \gamma\lambda_{iT} \\ \sum_{t=1}^{T}\mu_{it}\lambda_{it} & \sum_{t=1}^{T}\mu_{it}^2\lambda_{it} & \gamma\mu_{i1}\lambda_{i1} & \ldots & \gamma\mu_{iT}\lambda_{iT} \\ \gamma\lambda_{i1} & \gamma\mu_{i1}\lambda_{i1} & \gamma^2\lambda_{i1} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma\lambda_{iT} & \gamma\mu_{iT}\lambda_{iT} & 0 & \ldots & \gamma^2\lambda_{iT} \end{pmatrix}.$$

We can partition the previous matrix into

$$\mathbf{G}_i = \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{G}_{12}^{(i)} \\ \mathbf{G}_{12}^{(i)'} & \mathbf{G}_{22}^{(i)} \end{pmatrix}, \quad \text{(A10)}$$

Where $\mathbf{G}_{11}^{(i)}$ is the upper left-corner 2 × 2 submatrix of $\mathbf{G}_i$. Because there is no information in these cluster sizes about the μ's from other groups, the information about $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11},\ldots, \mu_{1T},\ldots, \mu_{m1},\ldots, \mu_{mT}, \sigma^2, \rho, \rho^*)'$ contained in the random cluster sizes for one subject in group i has the expression

$$\boldsymbol{I}_i^*(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{t=1}^{T}\lambda_{it} & \sum_{t=1}^{T}\mu_{it}\lambda_{it} & \mathbf{0}'_{(i-1)T} & \gamma\lambda_{i1} & \ldots & \gamma\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \sum_{t=1}^{T}\mu_{it}\lambda_{it} & \sum_{t=1}^{T}\mu_{it}^2\lambda_{it} & \mathbf{0}'_{(i-1)T} & \gamma\mu_{i1}\lambda_{i1} & \ldots & \gamma\mu_{iT}\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \mathbf{0}_{(i-1)T} & \mathbf{0}_{(i-1)T} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{0}_{(i-1)T} & \ldots & \mathbf{0}_{(i-1)T} & \mathbf{O}_{(i-1)T\times(m-i)T} & 0 & 0 & 0 \\ \gamma\lambda_{i1} & \gamma\mu_{i1}\lambda_{i1} & \mathbf{0}'_{(i-1)T} & \gamma^2\lambda_{i1} & \ldots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma\lambda_{iT} & \gamma\mu_{iT}\lambda_{iT} & \mathbf{0}'_{(i-1)T} & 0 & \ldots & \gamma^2\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \mathbf{0}_{(m-i)T} & \mathbf{0}_{(m-i)T} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{0}_{(m-i)T} & \ldots & \mathbf{0}_{(m-i)T} & \mathbf{O}_{(m-i)T\times(m-i)T} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \ldots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \ldots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \ldots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \end{pmatrix}$$

A simpler way of writing the previous matrix is

$$\boldsymbol{I}_i^*(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times(i-1)T} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times(m-i)T} & \mathbf{O}_{2\times3} \\ \mathbf{O}_{(i-1)T\times2} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{O}_{(i-1)T\times T} & \mathbf{O}_{(i-1)T\times(m-i)T} & \mathbf{O}_{(i-1)T\times3} \\ \mathbf{G}_{12}^{(i)'} & \mathbf{O}_{T\times(i-1)T} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times(m-i)T} & \mathbf{O}_{T\times3} \\ \mathbf{O}_{(m-i)T\times2} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{O}_{(m-i)T\times T} & \mathbf{O}_{(m-i)T\times(m-i)T} & \mathbf{O}_{(m-i)T\times3} \\ \mathbf{O}_{3\times2} & \mathbf{O}_{3\times(i-1)T} & \mathbf{O}_{3\times T} & \mathbf{O}_{3\times(m-i)T} & \mathbf{O}_{3\times3} \end{pmatrix}$$

Using the expression of $\boldsymbol{I}(\theta_i|\boldsymbol{k})$ from (A9), the information contributed by the clusters of severities given the clusters sizes, $\boldsymbol{I}_i(\theta|\boldsymbol{k})$, can be computed as

$$\boldsymbol{I}_i(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{2\times2} & \mathbf{O}_{2\times(i-1)T} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times(m-i)T} & \mathbf{O}_{2\times3} \\ \mathbf{O}_{(i-1)T\times2} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{O}_{(i-1)T\times T} & \mathbf{O}_{(i-1)T\times(m-i)T} & \mathbf{O}_{(i-1)T\times3} \\ \mathbf{O}_{T\times2} & \mathbf{O}_{T\times(i-1)T} & \frac{1}{\sigma^2}\sum_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times(m-i)T} & \mathbf{O}_{T\times3} \\ \mathbf{O}_{(m-i)T\times2} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{O}_{(m-i)T\times T} & \mathbf{O}_{(m-i)T\times(m-i)T} & \mathbf{O}_{(m-i)T\times3} \\ \mathbf{O}_{3\times2} & \mathbf{O}_{3\times(i-1)T} & \mathbf{O}_{3\times T} & \mathbf{O}_{3\times(m-i)T} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*) \end{pmatrix}$$

So, adding the corresponding pieces gives us

$$\boldsymbol{I}^{(i)}(\boldsymbol{\theta})=\boldsymbol{I}_i^*(\boldsymbol{\theta})+\sum_{k.\geq 1}\left[\prod_{t=1}^{T}e^{-\lambda_{it}}\frac{{\lambda_{it}}^{k_t}}{k_t!}\right]\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \quad \text{(A11)}$$

Let us now consider $n=\sum_{i=1}^{m}n_i$ independent observations from the multiple group model introduced in Section 3. Adding all the corresponding pieces gives us the information matrix for the multiple group model,

$$\boldsymbol{I}_n(\boldsymbol{\theta})=\sum_{i=1}^{m}n_i\boldsymbol{I}_{\boldsymbol{i}}^*(\boldsymbol{\theta})+\sum_{i=1}^{m}n_i\sum_{k.\geq 1}\left[\prod_{t=1}^{T}e^{-\lambda_{it}}\frac{{\lambda_{it}}^{k_t}}{k_t!}\right]\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \quad \text{(A12)}$$

It can easily be shown that the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about the parameter vector $\boldsymbol{\theta}$, contained in the $n=\sum_{i=1}^{m}n_i$ independent observations from the multiple group model, has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta})=\begin{pmatrix}\boldsymbol{I}_n(\delta,\gamma,\mu_{11},\ldots,\mu_{mT}) & \mathbf{O}_{(mT+2)\times 3}\\ \mathbf{O}_{3\times(mT+2)} & \sum_{i=1}^{m}n_i\sum_{k.\geq 1}\left[\prod_{t=1}^{T}e^{-\lambda_{it}}\frac{{\lambda_{it}}^{k_t}}{k_t!}\right]\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)\end{pmatrix}, \quad \text{(A13)}$$

where

$$\boldsymbol{I}_n(\delta,\gamma,\mu_{11},\ldots,\mu_{mT})=\sum_{i=1}^{m}n_i\begin{pmatrix}\mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times(i-1)T} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times(m-i)T}\\ \mathbf{O}_{(i-1)T\times 2} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{O}_{(i-1)T\times T} & \mathbf{O}_{(i-1)T\times(m-i)T}\\ \mathbf{G}_{12}^{(i)\prime} & \mathbf{O}_{T\times(i-1)T} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times(m-i)T}\\ \mathbf{O}_{(m-i)T\times 2} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{O}_{(m-i)T\times T} & \mathbf{O}_{(m-i)T\times(m-i)T}\end{pmatrix} \quad \text{(A14)}$$
$$+\frac{1}{\sigma^2}Diag\left(\mathbf{O}_{2\times 2},n_1\sum_{k.\geq 1}\prod_{t=1}^{T}e^{-\lambda_{1t}}\frac{{\lambda_{1t}}^{k_t}}{k_t!}\sum_{\boldsymbol{k}}^{-1},\ldots,n_m\sum_{k.\geq 1}\prod_{t=1}^{T}e^{-\lambda_{mt}}\frac{{\lambda_{mt}}^{k_t}}{k_t!}\sum_{\boldsymbol{k}}^{-1}\right.$$

The matrices $\mathbf{G}_{kl}^{(i)}$'s (k, l = 1,2) are given in (A10), and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are defined by (A3) – (A8).

References

1. Liang KY, Zeger SL. Longitudinal analysis using generalized linear models. Biometrika. 1986; 73(1):13–22.10.1093/biomet/73.1.13
2. Zeger SL, Liang KY. Longitudinal analysis for discrete and continuous outcomes. Biometrics. 1986; 42(1):121–130.10.2307/2531248 [PubMed: 3719049]
3. Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. Biometrika. 2001; 88(4):1121–1134.10.1093/biomet/88.4.1121
4. Williamson JM, Datta S, Satten GA. Marginal analyses of clustered data when cluster size is informative. Biometrics. 2003; 59(1):36–42.10.1111/1541-0420.00005 [PubMed: 12762439]

5. Wang M, Kong M, Datta S. Inference for marginal linear models for clustered longitudinal data with potentially informative cluster sizes. Statistical Methods in Medical Research. 2011; 20(4):347–367.10.1177/0962280209347043 [PubMed: 20223781]
6. Chiang CT, Lee KY. Efficient estimation methods for informative cluster size data. Statistica Sinica. 2008; 18:121–133.
7. Chen Z, Zhang B, Albert PS. A joint modeling approach to data with informative cluster size: robustness to the cluster size model. Statistics in Medicine. 2011; 30(15):1825–1836.10.1002/sim.4239 [PubMed: 21495060]
8. Follmann D, Proschan M, Leifer E. Multiple outputations: inference for complex clustered data by averaging analyses from independent data. Biometrics. 2003; 59(2):420–429.10.1111/1541-0420.00049 [PubMed: 12926727]
9. Follmann D, Fay M. Exact inference for complex clustered data using within-cluster resampling. Journal of Biopharmaceutical Statistics. 2010; 20(4):850–869.10.1080/10543401003618884 [PubMed: 20496210]
10. Cong XJ, Yin G, Shen Y. Marginal analyses of correlated failure time data with informative cluster sizes. Biometrics. 2007; 63(3):663–672.10.1111/j.1541-0420.2006.00730.x [PubMed: 17825000]
11. Williamson JM, Kim HY, Manatunga A, Addiss DG. Modeling survival data with informative cluster size. Statistics in Medicine. 2008; 27(4):543–555.10.1002/sim.3003 [PubMed: 17640035]
12. Barnhart HX, Sampson AR. Multiple population models for multivariate random length data without covariates. Biometrics. 1995; 51:195–204. [PubMed: 7766774]
13. Barnhart HX, Kosinski AS, Sampson AR. A regression model for multivariate random length data. Statistics in Medicine. 1999; 18(2):199–211.10.1002/(SICI)1097-0258(19990130)18:2<199::AID-SIM1>3.0.CO;2-E [PubMed: 10028140]
14. Allen AS, Barnhart HX. General marginal regression models for the joint modeling of event frequency and correlated severities with applications to clinical trials. Journal of Data Science. 2005; 3:199–219.
15. Williamson DE, Birmaher B, Frank E, Anderson BP, Matty MK, Kupfer DJ. Nature of life events and difficulties in depressed adolescents. Journal of the American Academy of Child and Adolescent Psychiatry. 1998; 37(10):1049–1057.10.1097/00004583-199810000-00015 [PubMed: 9785716]
16. Brown, GW.; Harris, TO. Life Events and Illness. Guilford Press; New York: 1989.
17. Monck E, Dobbs R. Measuring life events in an adolescent population: methodological issues and related findings. Psychological Medicine. 1985; 15(4):841–850. [PubMed: 4080887]
18. Hillegers MH, Burger H, Wals M, Reichart CG, Verhulst FC, Nolen WA, Ormel J. Impact of stressful life events, familial loading and their interaction on the onset of mood disorders. British Journal of Psychiatry. 2004; 185:97–101.10.1192/bjp.185.2.97 [PubMed: 15286059]
19. Rosnick CB, Small BJ, McEvoy CL, Borenstein AR, Mortimer JA. Negative life events and cognitive performance in a population of older adults. Journal of Aging and Health. 2007; 19(4):612–629.10.1177/0898264307300975 [PubMed: 17682077]
20. Kendler KS, Karkowski LM, Prescott CA. Causal relationship between stressful life events and the onset of major depression. American Journal of Psychiatry. 1999; 156(6):837–841. [PubMed: 10360120]
21. Lopez AD, Murray C. The global burden of disease, 1990–2020. Nature Medicine. 1998; 4(11):1241–1243.
22. Centers for Disease Control and Prevention. [accessed 20 July 2011] Depression in the United States household population. 2005. Available at http://www.cdc.gov/nchs/data/databriefs/db07.pdf
23. Barnhart, HX. Ph D Dissertation. University of Pittsburgh; Pittsburgh: 1992. Models for Multivariate Random Length Data with Applications in Clinical Trials.
24. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2012. http://www.R-project.org
25. Holgate P. Estimation for the bivariate Poisson distribution. Biometrika. 1964; 51:241–245.10.1093/biomet/51.1-2.241
26. McCulloch, CE.; Searle, SR.; Neuhaus, JM. Generalized, Linear, and Mixed Models. 2nd. Wiley; New York: 2008.

27. Iosif, AM. Ph D Dissertation. University of Pittsburgh; Pittsburgh: 2007. Analysis of longitudinal random length data.

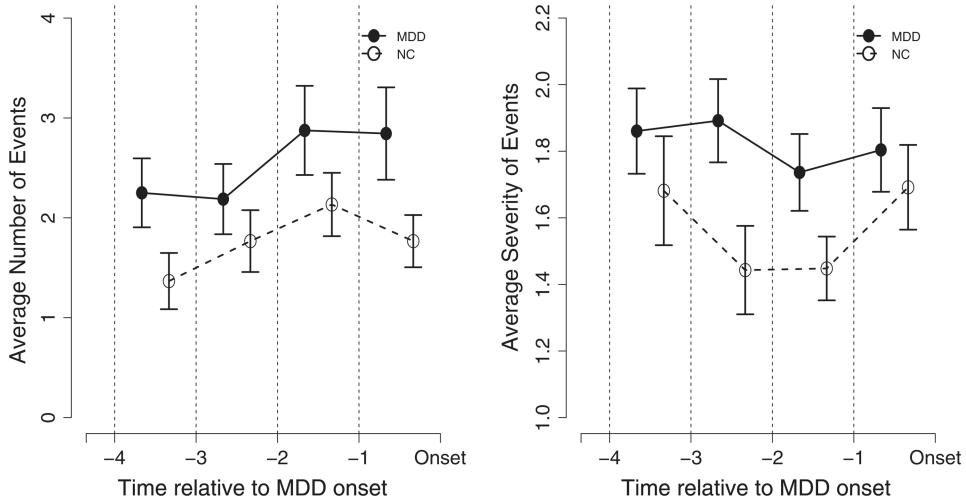

Figure 1.

Number and severity of stressful life events (mean ± standard error) for depressed (major depressive disorder (MDD)) and normal control (NC) adolescents. Events were recorded quarterly during the year before the onset of depression.

Table I

Choice of parameters for simulation study.

	$\boldsymbol{\mu}$'s similar and $\boldsymbol{\lambda}$'s disparate	$\boldsymbol{\mu}$'s disparate and $\boldsymbol{\lambda}$'s similar
$\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ parallel	$\boldsymbol{\mu_1}$ = (1.0, 1.1, 1.2, 1.3)	$\boldsymbol{\mu_1}$ = (1, 2, 3, 4)
	$\boldsymbol{\mu_2}$ = (1.5, 1.6, 1.7, 1.8)	$\boldsymbol{\mu}_2$ = (1.5, 2.5, 3.5, 4.5)
	$\boldsymbol{\lambda_1}$ = (2.7, 3.4, 4.3, 5.4)	$\boldsymbol{\lambda_1}$ = (12.3, 12.4, 12.6, 12.7)
	$\boldsymbol{\lambda_2}$ = (8.6, 10.8, 13.6, 17.1)	$\boldsymbol{\lambda}_2$ = (12.4, 12.5, 12.6, 12.7)
	$(\delta, \gamma) = (-1.3, 2.3)$	(δ, γ) (2.5, 0.01)
	$(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$	$(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$
μ_1 and μ_2 intersecting	$\boldsymbol{\mu_1}$ = (1.3, 1.2, 1.1, 1.0)	$\boldsymbol{\mu_1}$ = (4, 3, 2, 1)
	$\boldsymbol{\mu_2}$ = (1.0, 1.1, 1.2, 1.3)	$\boldsymbol{\mu}_2$ = (1, 2, 3, 4)
	$\boldsymbol{\lambda_1}$ = (9.5, 7.4, 5.8, 4.5)	$\boldsymbol{\lambda_1}$ = (5.5, 5.2, 5.0, 4.7)
	$\boldsymbol{\lambda_2}$ = (4.5, 5.8, 7.4, 9.5)	$\boldsymbol{\lambda}_2$ = (4.7, 5.0, 5.2, 5.5)
	$(\delta, \gamma) = (-1, 2.5)$	$(\delta, \gamma) = (1.5, 0.05)$
	$(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$	$(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$

Table II

Bias of the maximum likelihood estimator.

	Scenario 1				Scenario 2				Scenario 3				Scenario 4			
		Sample size				Sample size				Sample size				Sample size		
	True value	20	50	100	True value	20	50	100	True value	20	50	100	True value	20	50	100
μ_{11}	1.0	-0.005	-0.006	-0.002	1.3	-0.001	0.003	0.001	1.0	-0.006	0.006	-0.001	4	0.000	0.002	0.002
μ_{12}	1.1	-0.004	-0.006	-0.001	1.2	-0.001	0.001	0.001	2.0	-0.002	-0.001	-0.001	3	-0.005	0.000	-0.002
μ_{13}	1.2	-0.004	-0.003	-0.001	1.1	-0.006	0.002	0.001	3.0	-0.005	0.005	0.003	2	-0.005	0.000	-0.002
μ_{14}	1.3	-0.003	-0.005	-0.002	1.0	-0.007	0.001	0.002	4.0	0.000	0.000	-0.002	1	-0.007	-0.004	0.003
μ_{21}	1.5	0.003	-0.001	0.000	1.0	-0.007	0.002	0.002	1.5	0.009	0.009	0.002	1	0.000	0.002	0.002
μ_{22}	1.6	0.002	0.001	0.000	1.1	-0.004	0.001	0.001	2.5	0.009	0.011	0.001	2	0.001	-0.001	0.000
μ_{23}	1.7	0.004	0.000	0.001	1.2	-0.001	0.002	0.000	3.5	-0.003	0.004	0.004	3	-0.007	-0.003	0.000
μ_{24}	1.8	0.007	0.002	0.001	1.3	0.000	0.001	0.001	4.5	0.011	0.000	0.002	4	0.000	0.000	0.003
λ_{11}	2.7	0.009	0.000	0.000	9.5	-0.036	0.019	0.005	12.3	-0.001	0.001	-0.010	5.5	0.002	0.011	0.001
λ_{12}	3.4	0.019	-0.004	0.007	7.4	0.026	-0.009	0.007	12.4	-0.009	-0.001	-0.006	5.2	0.003	0.007	0.000
λ_{13}	4.3	0.003	0.019	0.001	5.8	-0.016	0.013	0.007	12.6	-0.013	0.000	-0.001	5.0	-0.002	0.007	0.001
λ_{14}	5.4	-0.003	-0.010	-0.007	4.5	-0.021	0.001	0.013	12.7	-0.011	0.002	0.003	4.7	0.004	0.006	0.003
λ_{21}	8.6	0.007	0.010	-0.007	4.5	-0.007	0.012	0.013	12.4	-0.004	0.002	-0.008	4.7	0.009	0.007	0.003
λ_{22}	10.8	-0.042	0.034	-0.012	5.8	0.012	0.000	0.012	12.5	-0.009	0.001	-0.004	5.0	0.001	0.007	0.002
λ_{23}	13.6	-0.025	-0.002	0.012	7.4	0.011	0.005	0.002	12.6	-0.014	0.003	0.001	5.2	-0.003	0.007	0.001
λ_{24}	17.1	0.008	0.033	-0.005	9.5	0.019	-0.020	0.010	12.7	-0.011	0.005	0.007	5.5	0.002	0.012	0.001
δ	-1.3	-0.369	-0.096	-0.040	-1.0	-0.517	-0.312	-0.156	2.5	0.000	0.000	0.000	1.5	-0.001	0.000	0.000
γ	2.3	0.255	0.071	0.028	2.5	0.444	0.267	0.133	0.01	0.000	0.000	0.000	0.05	0.000	0.000	0.000
σ^2	1.0	-0.033	-0.026	-0.023	1.0	-0.017	-0.012	-0.009	1.0	-0.029	-0.010	-0.007	1.0	-0.041	-0.023	-0.016
ρ	0.5	-0.019	-0.013	-0.011	0.5	-0.011	-0.007	-0.005	0.5	-0.018	-0.007	-0.004	0.5	-0.025	-0.013	-0.008
ρ^*	0.2	-0.02	-0.013	-0.010	0.2	-0.009	-0.007	-0.005	0.2	-0.008	-0.004	-0.001	0.2	-0.01	-0.008	-0.005

Table III

Empirical standard error of the estimators for severities for the proposed method and within-cluster resampling.

	Scenario 1				Scenario 2				Scenario 3				Scenario 4			
		SE				SE				SE				SE		
	True value	ML	WCR	RE*	True value	ML	WCR	RE	True value	ML	WCR	RE	True value	ML	WCR	RE
$n = 20$																
μ_{11}	1.0	0.160	0.204	1.64	1.3	0.109	0.166	2.31	1.0	0.165	0.171	1.07	4	0.167	0.177	1.13
μ_{12}	1.1	0.142	0.203	2.05	1.2	0.094	0.177	3.56	2.0	0.160	0.165	1.06	3	0.174	0.181	1.08
μ_{13}	1.2	0.122	0.189	2.41	1.1	0.099	0.187	3.58	3.0	0.161	0.163	1.02	2	0.177	0.180	1.03
μ_{14}	1.3	0.108	0.182	2.82	1.0	0.117	0.182	2.43	4.0	0.166	0.172	1.07	1	0.178	0.187	1.10
μ_{21}	1.5	0.100	0.173	3.01	1.0	0.116	0.189	2.63	1.5	0.169	0.174	1.06	1	0.180	0.193	1.16
μ_{22}	1.6	0.107	0.178	2.75	1.1	0.097	0.189	3.81	2.5	0.161	0.164	1.03	2	0.170	0.179	1.11
μ_{23}	1.7	0.120	0.163	1.84	1.2	0.094	0.182	3.76	3.5	0.165	0.168	1.04	3	0.174	0.183	1.10
μ_{24}	1.8	0.140	0.167	1.43	1.3	0.107	0.174	2.63	4.5	0.161	0.167	1.08	4	0.174	0.178	1.05
$n = 50$																
μ_{11}	1.0	0.098	0.130	1.74	1.3	0.071	0.108	2.32	1.0	0.104	0.108	1.07	4	0.113	0.117	1.06
μ_{12}	1.1	0.087	0.125	2.08	1.2	0.059	0.115	3.84	2.0	0.103	0.106	1.05	3	0.114	0.117	1.08
μ_{13}	1.2	0.075	0.125	2.76	1.1	0.060	0.113	3.58	3.0	0.104	0.105	1.02	2	0.114	0.117	1.05
μ_{14}	1.3	0.067	0.116	3.04	1.0	0.074	0.118	2.57	4.0	0.103	0.106	1.06	1	0.114	0.117	1.05
μ_{21}	1.5	0.063	0.112	3.21	1.0	0.071	0.118	2.74	1.5	0.102	0.104	1.05	1	0.112	0.119	1.13
μ_{22}	1.6	0.069	0.106	2.40	1.1	0.059	0.115	3.80	2.5	0.105	0.108	1.06	2	0.113	0.118	1.09
μ_{23}	1.7	0.079	0.112	2.00	1.2	0.057	0.116	4.05	3.5	0.101	0.104	1.06	3	0.112	0.117	1.09
μ_{24}	1.8	0.093	0.108	1.36	1.3	0.069	0.109	2.47	4.5	0.101	0.104	1.06	4	0.113	0.116	1.06
$n = 100$																
μ_{11}	1.0	0.068	0.091	1.80	1.3	0.051	0.079	2.42	1.0	0.073	0.075	1.06	4	0.077	0.080	1.07
μ_{12}	1.1	0.060	0.087	2.08	1.2	0.042	0.081	3.72	2.0	0.080	0.081	1.02	3	0.077	0.081	1.10
μ_{13}	1.2	0.052	0.084	2.64	1.1	0.043	0.081	3.50	3.0	0.074	0.077	1.07	2	0.076	0.079	1.08
μ_{14}	1.3	0.047	0.081	2.99	1.0	0.052	0.083	2.54	4.0	0.075	0.077	1.06	1	0.080	0.085	1.14
μ_{21}	1.5	0.042	0.076	3.22	1.0	0.050	0.080	2.54	1.5	0.071	0.072	1.02	1	0.078	0.082	1.11

	Scenario 1				Scenario 2				Scenario 3				Scenario 4			
		SE				SE				SE				SE		
	True value	ML	WCR	RE*	True value	ML	WCR	RE	True value	ML	WCR	RE	True value	ML	WCR	RE
μ_{22}	1.6	0.046	0.075	2.68	1.1	0.043	0.083	3.66	2.5	0.075	0.076	1.03	2	0.076	0.081	1.14
μ_{23}	1.7	0.053	0.074	1.92	1.2	0.043	0.080	3.57	3.5	0.073	0.075	1.03	3	0.077	0.079	1.08
μ_{24}	1.8	0.061	0.074	1.47	1.3	0.050	0.076	2.28	4.5	0.076	0.078	1.05	4	0.077	0.078	1.03

SE, standard error; WCR, within-cluster resampling; ML, maximum likelihood.

*Relative efficiency of the ML estimator with respect to the within-cluster resampling estimator is given by Var ($\hat{\mu wCR}$)/Var ($\hat{\mu ML}$).

Table IV

Coverage of the 95% confidence intervals.

	Scenario 1				Scenario 2				Scenario 3				Scenario 4			
		Sample size				Sample size				Sample size				Sample size		
	True value	20	50	100	True value	20	50	100	True value	20	50	100	True value	20	50	100
μ_{11}	1.0	91.8	94.8	93.8	1.3	93.7	95.6	94.3	1.0	94.1	95.5	94.8	4	96.0	93.3	95.4
μ_{12}	1.1	92.9	93.9	93.3	1.2	94.0	95.6	94.4	2.0	94.3	95.7	92.9	3	93.9	94.6	95.1
μ_{13}	1.2	92.9	93.2	93.6	1.1	94.5	93.9	94.2	3.0	95.0	94.8	94.7	2	94.0	94.2	95.3
μ_{14}	1.3	92.2	93.2	93.1	1.0	93.2	94.1	95.0	4.0	94.0	94.7	95.3	1	94.7	94.4	94.3
μ_{21}	1.5	91.8	93.3	94.4	1.0	93.5	94.1	95.3	1.5	93.8	94.6	95.7	1	93.2	95.6	94.5
μ_{22}	1.6	92.1	93.0	95.2	1.1	94.3	94.4	95.0	2.5	95.5	95.2	94.7	2	95.2	95.2	94.8
μ_{23}	1.7	93.0	92.9	94.1	1.2	94.4	94.5	93.7	3.5	93.9	94.8	94.7	3	93.7	94.5	94.8
μ_{24}	1.8	92.5	93.4	94.2	1.3	93.9	94.2	93.9	4.5	94.6	94.9	95.6	4	94.1	94.1	93.9
λ_{11}	2.7	95.1	94.5	94.6	9.5	95.0	95.2	95.9	12.3	94.8	93.5	95.3	5.5	95.4	95.5	95.3
λ_{12}	3.4	93.9	95.3	94.3	7.4	94.6	93.8	95.6	12.4	94.7	93.9	94.1	5.2	94.6	95.4	95.6
λ_{13}	4.3	94.5	95.4	95.1	5.8	94.8	94.5	96.7	12.6	94.2	94.3	95.1	5.0	95.5	95.0	95.7
λ_{14}	5.4	96.4	95.8	95.7	4.5	95.5	94.9	97.3	12.7	95.9	95.2	95.7	4.7	95.9	95.0	95.0
λ_{21}	8.6	95.6	94.1	94.1	4.5	95.1	94.5	94.9	12.4	94.6	93.2	94.8	4.7	95.6	95.8	95.4
λ_{22}	10.8	95.4	94.8	94.8	5.8	95.4	95.2	95.1	12.5	93.8	94.9	94.8	5.0	95.5	95.1	96.0
λ_{23}	13.6	93.3	94.7	94.9	7.4	94.2	95.0	94.4	12.6	95.3	94.7	95.1	5.2	95.1	95.9	95.1
λ_{24}	17.1	95.9	95.2	95.8	9.5	95.2	96.6	95.9	12.7	96.2	95.0	95.6	5.5	95.0	95.2	95.7
δ	−1.3	89.7	91.7	95.5	−1.0	88.1	92.2	91.5	2.5	94.0	94.2	95.0	1.5	95.0	95.0	94.5
γ	2.3	89.3	92.7	94.8	2.5	87.6	91.8	91.6	0.01	96.2	93.7	95.3	0.05	95.2	93.6	94.4
σ^2	1.0	88.5	88.3	87.5	1.0	91.6	92.3	93.8	1.0	87.8	92.3	93.1	1.0	85.8	90.9	92.1
ρ	0.5	92.1	92.1	91.0	0.5	94.4	94.8	94.0	0.5	90.2	92.6	94.1	0.5	89.9	92.2	92.1
ρ^*	0.2	90.8	93.0	94.3	0.2	93.0	94.3	94.2	0.2	90.9	93.1	94.2	0.2	93.1	93.2	94.4

Table V

Bias and coverage probability of the 95% CI of the maximum likelihood estimators for n=50 when incorrectly assuming a patterned exchangeable correlation for the severities.

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP
μ_{11}	1.0	-0.008	90.9	1.3	0.000	93.9	1	−0.002	94.0	4	0.000	94.9
μ_{12}	1.1	−0.005	91.7	1.2	0.000	94.9	2	−0.003	93.7	3	0.003	94.4
μ_{13}	1.2	−0.004	91.2	1.1	−0.002	94.8	3	−0.002	92.6	2	0.001	94.8
μ_{14}	1.3	−0.004	92.2	1.0	−0.002	94.2	4	−0.006	94.5	1	–0.003	92.5
μ_{21}	1.5	−0.002	93.8	1.0	−0.002	94.1	1.5	0.002	95.3	1	–0.006	95.6
μ_{22}	1.6	−0.002	95.2	1.1	−0.002	95.1	2.5	0.004	94.5	2	–0.001	93.9
μ_{23}	1.7	−0.002	95.9	1.2	−0.002	94.4	3.5	0.001	94.1	3	–0.005	95.3
μ_{24}	1.8	0.000	94.7	1.3	−0.001	94.3	4.5	0.005	94.6	4	–0.001	93.9
λ_{11}	2.7	−0.014	94.9	9.5	0.027	95.6	12.3	0.023	93.3	5.5	0.002	93.5
λ_{12}	3.4	0.004	94.7	7.4	0.014	95.4	12.4	0.015	93.6	5.2	–0.001	94.2
λ_{13}	4.3	0.010	93.9	5.8	−0.002	95.1	12.6	0.008	95.3	5.0	–0.003	95.2
λ_{14}	5.4	0.000	96.2	4.5	0.009	94.2	12.7	0.003	94.9	4.7	–0.003	96.1
λ_{21}	8.6	0.003	93.6	4.5	−0.002	94.5	12.4	0.019	93.4	4.7	–0.003	95.7
λ_{22}	10.8	0.010	92.4	5.8	0.005	95.1	12.5	0.012	94.4	5.0	–0.003	94.3
λ_{23}	13.6	−0.007	95.0	7.4	−0.006	95.5	12.6	0.006	94.8	5.2	–0.004	94.1
λ_{24}	17.1	0.023	93.4	9.5	−0.002	94.3	12.7	0.002	94.6	5.5	0.001	93.1
δ	−1.3	−0.083	92.5	−1.0	−0.338	92.8	2.5	0.002	94.1	1.5	–0.002	95.6
γ	2.3	0.060	92.9	2.5	0.291	92.5	0.01	0.000	94.8	0.05	0.000	95.2

True correlation is 0.5 for the severities within a measurement occasion and AR (0.2) for severities from different measurement occasions.

CP, coverage probability.

Table VI

Bias and coverage probability of the 95% CI of the maximum likelihood estimators for $n = 50$ when incorrectly assuming a patterned exchangeable correlation for the severities.

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP
μ_{11}	1.0	−0.004	91.9	1.3	−0.001	90.2	1	0.009	95.1	4	0.003	95.2
μ_{12}	1.1	−0.004	92.1	1.2	−0.001	92.3	2	0.001	94.7	3	−0.001	93.4
μ_{13}	1.2	−0.004	93.3	1.2	0.001	92.4	3	−0.006	93.2	2	−0.003	95.0
μ_{14}	1.3	−0.001	92.9	1.0	0.000	92.4	4	0.003	93.3	1	−0.001	94.3
μ_{21}	1.5	−0.002	93.1	1.0	0.002	92.2	1.5	0.000	94.8	1	−0.009	92.7
μ_{22}	1.6	0.003	92.7	1.1	0.001	93.2	2.5	0.000	94.6	2	−0.002	94.3
μ_{23}	1.7	0.003	92.1	1.2	−0.001	93.0	3.5	0.005	95.4	3	0.002	93.6
μ_{24}	1.8	0.004	91.6	1.3	−0.002	90.5	4.5	0.004	94.4	4	0.003	94.9
λ_{11}	2.7	0.021	94.7	9.5	0.008	95.0	12.3	−0.009	95.0	5.5	0.014	95.2
λ_{12}	3.4	0.016	95.7	7.4	−0.018	94.7	12.4	−0.005	95.0	5.2	0.005	95.3
λ_{13}	4.3	0.006	96.3	5.8	0.006	95.5	12.6	0.001	95.8	5.0	0.000	95.4
λ_{14}	5.4	0.021	95.4	4.5	−0.007	94.1	12.7	0.012	94.7	4.7	−0.003	95.2
λ_{21}	8.6	−0.033	94.3	4.5	0.000	95.5	12.4	−0.009	94.8	4.7	−0.005	94.6
λ_{22}	10.8	0.022	94.9	5.8	0.008	94.9	12.5	−0.002	95.5	5.0	0.000	95.9
λ_{23}	13.6	−0.015	95.3	7.4	−0.007	94.6	12.6	0.006	95.3	5.2	0.006	95.4
λ_{24}	17.1	−0.021	94.9	9.5	−0.009	95.7	12.7	0.018	94.1	5.5	0.014	95.6
δ	−1.3	−0.080	90.8	−1.0	−0.719	87.6	2.5	−0.002	95.3	1.5	−0.003	94.1
γ	2.3	0.056	90.0	2.5	0.624	87.2	0.01	0.001	94.9	0.05	0.001	94.6

True correlation is 0.2 for the severities from different measurement occasions and 0.2 + AR (0.5) for severities within a measurement occasion.

CP, coverage probability.

Table VII

Bias and coverage probability of the 95% CI of the maximum likelihood estimators for $n = 50$ when incorrectly assuming independent Poisson distributions for the cluster sizes.

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP
μ_{11}	1.0	−0.003	94.6	1.3	0.002	95.5	1.0	0.001	94.4	4	0.001	94.9
μ_{12}	1.1	−0.003	93.4	1.2	0.001	94.8	2.0	−0.005	95.6	3	−0.003	95.6
μ_{13}	1.2	−0.004	93.8	1.1	0.001	94.8	3.0	−0.008	95.2	2	0.002	93.8
μ_{14}	1.3	−0.003	93.7	1.0	0.001	93.8	4.0	−0.001	94.4	1	−0.002	93.9
μ_{21}	1.5	−0.003	92.8	1.0	−0.001	94.3	1.5	0.000	95.6	1	0.005	93.4
μ_{22}	1.6	−0.003	93.8	1.1	−0.001	94.1	2.5	0.002	95.0	2	0.003	91.9
μ_{23}	1.7	−0.003	94.5	1.2	0.000	95.3	3.5	0.008	94.0	3	0.002	94.1
μ_{24}	1.8	−0.003	94.2	1.3	0.000	95.0	4.5	0.002	96.3	4	0.000	94.3
λ_{11}	2.7	0.002	95.3	9.5	0.016	94.5	12.3	0.000	96.1	5.5	0.000	93.9
λ_{12}	3.4	0.001	95.3	7.4	−0.003	94.7	12.4	−0.009	94.7	5.2	−0.004	89.2
λ_{13}	4.3	−0.012	93.9	5.8	0.007	95.4	12.6	−0.017	92.5	5.0	−0.004	90.2
λ_{14}	5.4	0.012	95.9	4.5	0.003	95.2	12.7	−0.023	93.9	4.7	−0.005	92.9
λ_{21}	8.6	0.001	94.2	4.5	0.015	94.6	12.4	−0.005	95.5	4.7	−0.004	92.5
λ_{22}	10.8	−0.002	95.6	5.8	−0.017	94.0	12.5	−0.013	93.7	5.0	−0.004	90.0
λ_{23}	13.6	−0.005	96.0	7.4	−0.008	94.6	12.6	−0.019	92.8	5.2	−0.003	90.8
λ_{24}	17.1	0.019	94.9	9.5	−0.023	94.6	12.7	−0.024	94.6	5.5	0.000	93.9
δ	−1.3	−0.116	93.0	−1.0	−0.354	91.5	2.5	0.000	96.4	1.5	−0.002	94.7
γ	2.3	0.085	92.8	2.5	0.301	92.1	0.01	−0.001	96.3	0.05	0.000	96.9

True model for the cluster sizes is multivariate Poisson.

CP, coverage probability.

Stat Med. Author manuscript; available in PMC 2015 February 28.

Table VIII

Bias and coverage probability of the 95% CI of the maximum likelihood estimators for $n = 50$ when incorrectly assuming a Poisson distribution for the cluster sizes.

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP	True value	Bias	CP
μ_{11}	1.0	−0.003	91.6	1.3	−0.002	92.2	1.0	−0.003	94.7	4	0.002	93.6
μ_{12}	1.1	0.001	92.2	1.2	0.001	91.7	2.0	−0.006	93.5	3	−0.004	92.0
μ_{13}	1.2	−0.002	92.0	1.1	0.001	93.5	3.0	−0.003	94.1	2	0.006	92.6
μ_{14}	1.3	−0.004	90.8	1.0	0.002	92.3	4.0	−0.001	94.2	1	0.002	93.1
μ_{21}	1.5	−0.002	93.6	1.0	0.002	93.3	1.5	0.005	95.2	1	−0.003	91.8
μ_{22}	1.6	−0.005	93.2	1.1	0.001	92.6	2.5	0.003	95.1	2	−0.003	95.7
μ_{23}	1.7	−0.004	93.2	1.2	0.001	92.4	3.5	−0.004	93.9	3	− 0.004	93.4
μ_{24}	1.8	−0.006	94.0	1.3	−0.003	92.8	4.5	0.004	92.8	4	0.004	91.7
λ_{11}	2.7	−0.026	90.4	9.5	0.000	84.1	12.3	0.006	79.5	5.5	−0.002	89.5
λ_{12}	3.4	0.007	89.8	7.4	0.031	85.8	12.4	−0.004	81.5	5.2	−0.001	90.1
λ_{13}	4.3	−0.007	88.0	5.8	−0.009	89.8	12.6	−0.007	81.1	5.0	0.000	89.4
λ_{14}	5.4	−0.024	89.2	4.5	−0.006	88.8	12.7	0.000	80.5	4.7	0.003	90.0
λ_{21}	8.6	0.013	87.4	4.5	−0.015	90.2	12.4	−0.007	80.7	4.7	−0.001	89.3
λ_{22}	10.8	−0.022	83.8	5.8	−0.002	87.9	12.5	0.012	82.5	5.0	−0.003	90.0
λ_{23}	13.6	0.039	79.4	7.4	0.028	86.5	12.6	−0.008	81.6	5.2	−0.001	90.1
λ_{24}	17.1	0.002	81.0	9.5	−0.018	83.9	12.7	−0.006	80.6	5.5	0.005	90.3
δ	−1.3	−0.146	92.8	−1.0	−0.382	91.5	2.5	0.000	80.1	1.5	−0.002	88.9
γ	2.3	0.105	93.0	2.5	0.327	90.5	0.01	0.000	80.8	0.05	0.000	88.7

True model for the cluster sizes is $NB(10, e^{\delta+\gamma\mu_{it}})$.

CP, coverage probability.

Table IX

Life events and difficulties schedule data.

	Number of stressful life events experienced										
	0	1	2	3	4	5	6	7	9	11	Total
First 3-month period before MDD onset											
MDD	6	8	2	3	7	1	2	2	—	1	91
NC	6	9	6	6	1	2	—	—	—	—	53
Second 3-month period before MDD onset											
MDD	6	4	7	4	4	3	1	2	—	1	92
NC	6	8	3	6	4	2	1	—	—	—	64
Third 3-month period before MDD onset											
MDD	8	5	8	5	—	2	4	—	—	—	70
NC	11	3	4	9	1	1	1	—	—	—	53
Fourth 3-month period before MDD onset											
MDD	5	9	5	7	2	3	—	—	1	—	72
NC	13	4	6	5	1	—	1	—	—	—	41

Frequency of life events for the 32 depressed (MDD) and 30 normal control (NC) adolescents. MDD, major depressive disorder.

Table X

Maximum likelihood estimates for life events and difficulties schedule data.

	Parameter	$\hat{\theta}$	Estimated SD of $\hat{\theta}$	Initial value
	δ	−16:6291	27.91	−0:1986
	γ	9.4031	15.08	0.5683
MDD	μ_{11}	1.8563	0.04	1.8605
	μ_{12}	1.8515	0.04	1.8917
	μ_{13}	1.8804	0.06	1.7364
	μ_{14}	1.8799	0.06	1.8039
	λ_{11}	2.2848	0.27	2.3599
	λ_{12}	2.1839	0.26	2.4021
	λ_{13}	2.8653	0.30	2.1992
	λ_{14}	2.8524	0.30	2.2853
NC	μ_{21}	1.8034	0.09	1.6814
	μ_{22}	1.8271	0.06	1.4430
	μ_{23}	1.8481	0.04	1.4479
	μ_{24}	1.8287	0.06	1.6917
	λ_{21}	1.3891	0.21	2.1315
	λ_{22}	1.7351	0.24	1.8615
	λ_{23}	2.1148	0.26	1.8667
	λ_{24}	1.7622	0.24	2.1440
	σ^2	0.5250	0.03	0.5324
	ρ	0.1702	0.05	0.1846
	ρ^*	0.0689	0.04	0.0787

MDD, major depressive disorder; NC, normal control; SD, standard deviation.