Modeling Individual Patient Count/Rate Data over Time with Applications to Cancer Pain Flares and Cancer Pain Medication Usage

George J. Knafl[1*], Salimah H. Meghani[2]

[1]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Department of Biobehavioral Health Sciences, School of Nursing, University of Pennsylvania, Philadelphia, PA, USA
Email: *gknafl@unc.edu, meghagnis@nursing.upenn.edu

Abstract

The purpose of this article is to investigate approaches for modeling individual patient count/rate data over time accounting for temporal correlation and non-constant dispersions while requiring reasonable amounts of time to search over alternative models for those data. This research addresses formulations for two approaches for extending generalized estimating equations (GEE) modeling. These approaches use a likelihood-like function based on the multivariate normal density. The first approach augments standard GEE equations to include equations for estimation of dispersion parameters. The second approach is based on estimating equations determined by partial derivatives of the likelihood-like function with respect to all model parameters and so extends linear mixed modeling. Three correlation structures are considered including independent, exchangeable, and spatial autoregressive of order 1 correlations. The likelihood-like function is used to formulate a likelihood-like cross-validation (LCV) score for use in evaluating models. Example analyses are presented using these two modeling approaches applied to three data sets of counts/rates over time for individual cancer patients including pain flares per day, as needed pain medications taken per day, and around the clock pain medications taken per day per dose. Means and dispersions are modeled as possibly nonlinear functions of time using adaptive regression modeling methods to search through alternative models compared using LCV scores. The results of these analyses demonstrate that extended linear mixed modeling is preferable for modeling individual patient count/rate data over time, because in example analyses, it either generates better LCV scores or more parsimonious models and requires substantially less time.

Keywords

Adaptive Regression, Extended Linear Mixed Modeling, Generalized

1. Introduction

An ongoing study (NIH/NINR 1R01NR017853) of patients with cancer is collecting daily longitudinal count/rate data including numbers of pain flares per day and numbers of as needed pain medications taken per day. Data are being collected for each study participant over periods of up to five months long. A completed study (NIH/NINR RC1NR011591) collected numbers for cancer patients over 3 months of around the clock pain medications taken per day per dose, that is, the number of times a medication is taken in a day relative to the number of doses that are supposed to be taken in a day. Standard assumptions of means linear in time and dispersions constant over time are not always appropriate for such data. Also, a Poisson process assumption of independence over time needs not always hold. A model selection score needs to be defined for evaluating models for the data and for use in searches over alternative models. Times to conduct these searches need to be as short as possible, especially as the number of time measurements increases.

Approaches are presented for modeling mean counts/rates over time separately for each individual patient controlling for temporal correlation as well as for time-varying dispersions. These approaches use Poisson regression methods, because count/rate data are being modeled. Generalized estimating equations (GEE) methods [1] [2] provide a natural choice for modeling correlations for such count variables. However, standard GEE methods have limited value, because they assume constant dispersions. Furthermore, GEE methods avoid specification of likelihood functions, which are useful for generating model selection criteria. In what follows, two extensions of GEE methods are formulated and evaluated that address temporal correlation and time-varying means and dispersions for repeated count/rate measurements. A likelihood-like function, that is, a function used like a likelihood but which needs not integrate to 1, is defined and used in computing parameter estimates for these extensions along with a model selection criterion for comparing alternative models. Example analyses of selected individual cancer patient count/rate data are presented using adaptive regression methods [3] for identifying possibly nonlinear trajectories for means and dispersions of counts/rates over time while controlling for temporal correlation.

2. Modeling Individual Count/Rate Data

2.1. Generalized Linear Modeling of Means

Let $y_{t(i)}$ denote count values for an individual patient observed at N distinct times within a general set T of times, that is, $t(i) \in T = \{t(i) : 1 \leq i \leq N\}$. Combine these into the $N \times 1$ vector $\boldsymbol{y}$. Let $\mu_{t(i)} = E y_{t(i)}$ denote associated mean

or expected counts and combine these into the $N \times 1$ vector $\boldsymbol{\mu}$. Denote the residuals as $e_{t(i)} = y_{t(i)} - \mu_{t(i)}$ for $t(i) \in T$ and combine these into the $N \times 1$ vector $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{\mu}$. Let $x_{t(i),j}$ denote predictor values over times $t(i) \in T$ and over predictors indexed by $1 \leq j \leq J$ and combine these into the $J \times 1$ vector $\boldsymbol{x}_{t(i)}$ with transpose denoted by $\boldsymbol{x}_{t(i)}^{\mathrm{T}}$ for $t(i) \in T$. Let $\mathbf{X}$ be the $N \times J$ matrix with rows $\boldsymbol{x}_{t(i)}^{\mathrm{T}}$ for $1 \leq i \leq N$. Let $\boldsymbol{\beta}$ denote the associated $J \times 1$ vector of coefficient parameters. Use generalized linear models [4] [5] of $\mu_{t(i)}$ for $t(i) \in T$ with natural log link function $h(\mu) = \log_e(\mu)$ so that $h\left(\mu_{t(i)}\right) = \boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta}$ for $t(i) \in T$. When $x_{t(i),1} = 1$ for $t(i) \in T$, the first entry β_1 of $\boldsymbol{\beta}$ is an intercept parameter. Treat each $y_{t(i)}$ as Poisson distributed so that its variance is $V\left(\mu_{t(i)}\right) = \mu_{t(i)}$.

The counts $y_{t(i)}$ sometimes have associated totals $Y_{t(i)} > 0$, and then the model for the mean counts $\mu_{t(i)}$ is converted to a model for the means $\mu'_{t(i)}$ of the rates $y'_{t(i)} = y_{t(i)} / Y_{t(i)}$ using offsets $o_{t(i)} = \log\left(Y_{t(i)}\right)$. Formally, replace $\boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta}$ by $\boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta} + o_{t(i)}$ so that mean counts are $\mu_{t(i)} = \exp\left(\boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta} + o_{t(i)}\right)$ and then

$$\mu'_{t(i)} = E y'_{t(i)} = \mu_{t(i)} / Y_{t(i)} = \exp\left(\boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta}\right)$$

are the mean rates.

2.2. Time-Varying Dispersions

Let $x'_{t(i),j}$ denote predictor values over times $t(i) \in T$ and over predictors indexed by $1 \leq j \leq J'$ and combine these into the $J' \times 1$ vectors $\boldsymbol{x}'_{t(i)}$ for $t(i) \in T$. Let $\mathbf{X}'$ be the $N \times J'$ matrix with rows $\boldsymbol{x}'^{\mathrm{T}}_{t(i)}$ for $1 \leq i \leq N$. Let $\boldsymbol{\beta}'$ denote the associated $J' \times 1$ vector of coefficient parameters. Let $\varphi_{t(i)}$ denote dispersion values over times $t(i) \in T$ satisfying $\log\left(\varphi_{t(i)}\right) = \boldsymbol{x}'^{\mathrm{T}}_{t(i)} \cdot \boldsymbol{\beta}'$ and define the extended variances as

$$\sigma^2_{t(i)} = \varphi_{t(i)} \cdot V\left(\mu_{t(i)}\right) = \varphi_{t(i)} \cdot \mu_{t(i)}$$

and the extended standard deviations as $\sigma_{t(i)} = \varphi^{1/2}_{t(i)} \cdot \mu^{1/2}_{t(i)}$ for $t(i) \in T$. Informally, these quantities extend the usual Poisson variances and standard deviations through multiplication by dispersions. These are used to compute the standardized residuals $stde_{t(i)} = e_{t(i)} / \sigma_{t(i)}$. Combine the extended standard deviations into the $N \times 1$ vector $\boldsymbol{\sigma}$. When $x'_{t(i),1} = 1$ for $t(i) \in T$, the first entry β'_1 of $\boldsymbol{\beta}'$ is an intercept parameter. The constant dispersion model corresponds to $x'_{t(i),1} = 1$ for $t(i) \in T$ with $J' = 1$. This is the dispersion model used in standard GEE modeling.

When offsets $o_{t(i)}$ are used to convert the model for the counts $y_{t(i)}$ to a model for the rates $y'_{t(i)}$, they can also be added to the dispersions. The dispersions then satisfy $\log\left(\varphi_{t(i)}\right) = \boldsymbol{x}'^{\mathrm{T}}_{t(i)} \cdot \boldsymbol{\beta}' + o_{t(i)}$ so that the extended variances for the counts $y_{t(i)}$ are

$$\sigma^2_{t(i)} = \varphi_{t(i)} \cdot \mu_{t(i)} = \exp\left(\boldsymbol{x}'^{\mathrm{T}}_{t(i)} \cdot \boldsymbol{\beta}'\right) \cdot \exp\left(\boldsymbol{x}_{t(i)}^{\mathrm{T}} \cdot \boldsymbol{\beta}\right) \cdot \exp^2\left(o_{t(i)}\right)$$

and then the variances for the rates $y'_{t(i)}$ are

$$\sigma'^2_{t(i)} = \sigma^2_{t(i)} \big/ T^2_{t(i)} = \varphi'_{t(i)} \cdot \mu'_{t(i)}$$

where

$$\varphi'_{t(i)} = \exp\left(\boldsymbol{x}'^{\mathrm{T}}_{t(i)} \cdot \boldsymbol{\beta}'\right).$$

2.3. Modeling Correlations

Denote the covariance matrix for the count vector $\boldsymbol{y}$ as $\boldsymbol{\Sigma}$. Use the GEE approach [1] [2] to model the covariance matrix for $\boldsymbol{y}$ as $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\sigma}) \cdot \mathbf{R}(\rho) \cdot \mathrm{Diag}(\boldsymbol{\sigma})$ where $\mathrm{Diag}(\boldsymbol{\sigma})$ is the $N \times N$ diagonal matrix with diagonal entries $\sigma_{t(i)}$ for $t(i) \in T$ and $\mathbf{R}(\rho)$ is a $N \times N$ correlation matrix with diagonal entries 1 and off diagonal entries $R_{t(i),t(i')}$ for $1 \le i \ne i' \le N$ determined by a correlation parameter ρ varying with the assumed correlation structure. Under independent (IND) correlations, $R_{t(i),t(i')} = \rho_{\mathrm{IND}} = 0$ for $1 \le i \ne i' \le N$. A Poisson process generates such correlations. Under exchangeable (EXCH) correlations, $R_{t(i),t(i')} = \rho_{\mathrm{EXCH}}$ for $1 \le i \ne i' \le N$, that is, the correlations are constant. Under autoregressive of order 1 (AR1) correlations, $R_{t(i),t(i')} = \rho_{\mathrm{AR1}}^{|t(i)-t(i')|}$ for $1 \le i \ne i' \le N$ where $|t(i)-t(i')|$ is the absolute value of $t(i)-t(i')$. These differences are all assumed to be integers so that the correlations $R_{t(i),t(i')}$ are all well-defined. In general, $R_{t(i),t(i')}$ are spatial AR1 correlations. The special case of non-spatial AR1 correlations with $t(i) = i$ treats times as equally spaced. The parameter ρ_{AR1} is called the autocorrelation.

2.4. Possible Extensions

The above formulation can be extended to address repeated measurements of types other than counts/rates and for multiple patients. More complex correlation structures based on multiple correlation parameters can also be considered. One such example is unstructured correlations with different correlations for different pairs of measurements, but this requires data from multiple patients to be reasonably estimated. These extensions are not addressed further.

3. Standard Generalized Estimating Equations Modeling

3.1. Notation and Parameter Estimation

Under standard GEE modeling, dispersions are treated as a constant φ_0 so that the covariance matrix satisfies

$$\boldsymbol{\Sigma} = \varphi_0 \cdot \mathrm{Diag}\left(V^{1/2}(\boldsymbol{\mu})\right) \cdot \mathbf{R}(\rho) \cdot \mathrm{Diag}\left(V^{1/2}(\boldsymbol{\mu})\right)$$

where $V(\boldsymbol{\mu})$ is the $N \times 1$ vector with entries $V\left(\mu_{t(i)}\right) = \mu_{t(i)}$ for $t(i) \in T$. The generalized estimating equations are given by $\boldsymbol{g}(\boldsymbol{\beta}) = \mathbf{0}$ where $\mathbf{0}$ is the $J \times 1$ vector with all zero entries, $\boldsymbol{g}(\boldsymbol{\beta}) = \mathbf{D}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{e}$, and the $N \times J$ matrix $\mathbf{D} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$ with entries $D_{t(i),j} = \partial\mu_{t(i)} \big/ \partial\beta_j = x_{t(i),j} \cdot \mu_{t(i)}$ for $t(i) \in T$ and $1 \le j \le J$. Let $\mathbf{H}(\boldsymbol{\beta}) = -\mathbf{D}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{D}$. Note that in the general GEE context with correlated outcomes for multiple subjects, the formulation for $\boldsymbol{g}(\boldsymbol{\beta})$ would equal a sum of terms like $\mathbf{D}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{e}$ for each subject and $\mathbf{H}(\boldsymbol{\beta})$ would equal

a sum of terms like $-\mathbf{D}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{D}$ for each subject. Only one such term is needed here since data for only one subject/patient are being modeled. The GEE process for estimating $\boldsymbol{\beta}$ iteratively solves $\mathbf{g}(\boldsymbol{\beta}) = 0$ as follows. Given the current value $\boldsymbol{\beta}_u$ for $\boldsymbol{\beta}$, the next value is given by $\boldsymbol{\beta}_{u+1} = \boldsymbol{\beta}_u - \mathbf{H}^{-1}(\boldsymbol{\beta}_u) \cdot \mathbf{g}(\boldsymbol{\beta}_u)$, thereby adapting Newton's method with $\mathbf{g}(\boldsymbol{\beta})$ in the role of the gradient vector and $\mathbf{H}(\boldsymbol{\beta})$ in the role of the Hessian matrix.

The constant dispersion parameter φ_0 is estimated using the Pearson residuals $Pe_{t(i)}(\boldsymbol{\beta}) = e_{t(i)} / V^{1/2}(\mu_{t(i)})$ evaluated at a given value for the mean coefficient parameter vector $\boldsymbol{\beta}$. The bias-adjusted estimate $\varphi_0(\boldsymbol{\beta})$ of the dispersion parameter φ_0 satisfies $\varphi_0(\boldsymbol{\beta}) = \sum_{i=1}^{N} Pe_{t(i)}^2(\boldsymbol{\beta}) / (N-J)$ assuming $N-J > 0$. Next, the correlation parameter $\rho(\boldsymbol{\beta})$ is estimated using standardized errors

$$stde_{t(i)}(\boldsymbol{\beta}) = \varphi_0^{-1/2}(\boldsymbol{\beta}) \cdot Pe_{t(i)}(\boldsymbol{\beta})$$

for $t(i) \in T$ as follows. The IND correlation structure has no need for an estimate. For the EXCH correlation structure and a given value $\boldsymbol{\beta}$ for the mean parameter vector, ρ_{EXCH} can be estimated by

$$\rho_{\mathrm{EXCH}}(\boldsymbol{\beta}) = \sum_{i=1}^{N-1} \sum_{i'=i+1}^{N} stde_{t(i)}(\boldsymbol{\beta}) \cdot stde_{t(i')}(\boldsymbol{\beta}) / (N \cdot (N-1)/2 - J)$$

assuming $N \cdot (N-1)/2 - J > 0$. For the AR1 correlation structure and a given value $\boldsymbol{\beta}$ for the mean parameter vector, the autocorrelation ρ_{AR1} can be estimated by

$$\rho_{\mathrm{AR1}}(\boldsymbol{\beta}) = \sum_{i=1}^{N-1} \left(stde_{t(i)}(\boldsymbol{\beta}) \cdot stde_{t(i+1)}(\boldsymbol{\beta}) \right)^{1/(|t(i)-t(i+1)|)} / (N-1-J)$$

assuming $N-1-J > 0$. In the non-spatial AR1 special case,

$$\rho_{\mathrm{AR1}}(\boldsymbol{\beta}) = \sum_{i=1}^{N-1} \left(stde_{t(i)}(\boldsymbol{\beta}) \cdot stde_{t(i+1)}(\boldsymbol{\beta}) \right) / (N-1-J)$$

because $|t(i) - t(i+1)| = 1$ for $1 \leq i < N$.

For any correlation structure, once the GEE estimate $\boldsymbol{\beta}(T)$ of the coefficient parameter vector $\boldsymbol{\beta}$ is computed using the observations indexed by $t(i) \in T$, the GEE estimate of the dispersion parameter φ_0 is $\varphi_0(T) = \varphi_0(\boldsymbol{\beta}(T))$. The GEE estimate of the correlation parameter ρ is $\rho(T) = \rho(\boldsymbol{\beta}(T))$ computed using $\boldsymbol{\beta}(T)$ and $\varphi_0(T)$.

3.2. The Likelihood-Like Function

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}} \varphi_0)^{\mathrm{T}}$ be the $(J+1) \times 1$ vector of the GEE mean and dispersion parameters. The correlation parameter ρ is a function of $\boldsymbol{\beta}$ and φ_0 and so has not been included in $\boldsymbol{\theta}$. Use the multivariate normal likelihood to define the likelihood-like function $L(T; \boldsymbol{\theta})$ satisfying

$$\ell(T; \boldsymbol{\theta}) = \log(L(T; \boldsymbol{\theta})) = -e^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot e/2 - \log(|\boldsymbol{\Sigma}|)/2 - N \cdot \log(2 \cdot \pi)/2$$

where $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix $\boldsymbol{\Sigma}$. The vector $\partial \ell(T; \boldsymbol{\theta}) / \partial \boldsymbol{\beta}$ of partial derivatives of $\ell(T; \boldsymbol{\theta})$ can be expressed as the sum of two terms. The first term corresponds to differentiating the residual vector part e of $\ell(T; \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ holding the covariance part $\boldsymbol{\Sigma}$ fixed in $\boldsymbol{\beta}$

and equals $g(\boldsymbol{\beta})$, the gradient-like quantity used in standard GEE modeling. This fact seems to have been first recognized by Chaganty [6]. One advantage for having a likelihood-like function for GEE models is that it can be used to compute parameter estimates. Another is that it can be used to compute model selection criteria not otherwise available for GEE modeling.

3.3. Likelihood-Like Cross-Validation

Burman [7] defined k-fold cross-validation with observations partitioned into k disjoint subsets called folds. Fold observations are predicted using parameter estimates computed using the data from the other folds. In k-fold likelihood-like cross-validation (LCV), these deleted fold predictions are scored using the associated likelihood-like function L. Randomly partition the times $t(i) \in T$ into k disjoint folds $T(f)$ for $1 \le f \le k$. Use the same initial seed for randomization with all models under consideration so that their LCV scores are comparable. Let $\boldsymbol{\theta}(T \setminus T(f))$ denote the estimate of $\boldsymbol{\theta}$ using the data with times in the complement $T \setminus T(f)$ of the fold $T(f)$. For $1 \le f \le k$, let $T^+(f)$ denote the union of all folds $T(f')$ for $1 \le f' \le f$ with $T^+(0)$ the empty fold and set $L(T^+(0); \boldsymbol{\theta}) = 1$. Define the LCV score to satisfy

$$\mathrm{LCV} = \prod_{f=1}^{k} \mathrm{LCV}_f^{1/N}$$

where LCV_f is defined as the conditional likelihood-like term for the data in fold $T(f)$ conditioned on the data in the union $T^+(f-1)$ of the prior folds using the deleted estimate $\boldsymbol{\theta}(T \setminus T(f))$ of the parameter vector $\boldsymbol{\theta}$. Formally,

$$\begin{aligned}
\mathrm{LCV}_f &= L\big(T(f) \,|\, T^+(f-1); \boldsymbol{\theta}(T \setminus T(f))\big) \\
&= L\big(T^+(f); \boldsymbol{\theta}(T \setminus T(f))\big) \big/ L\big(T^+(f-1); \boldsymbol{\theta}(T \setminus T(f))\big)
\end{aligned}$$

Because fold assignment is random, folds can be empty when the number k of folds is large relative to the number N of measurements, and then those folds are dropped from the computation of the LCV score. Larger LCV scores indicate better models. Note that even if the full data are non-spatial with observations at consecutive integer times $t(i) = i$ for $1 \le i \le N$, the folds $T(f)$ and the fold unions $T^+(f)$ are not consecutive integer times except in rare cases and so require more general handling.

4. Incorporating Nonconstant Dispersions

4.1. Formulation

GEE modeling can be extended to handle nonconstant dispersions. Let $\boldsymbol{\theta} = \big(\boldsymbol{\beta}^\mathrm{T} \boldsymbol{\beta}'^\mathrm{T}\big)^\mathrm{T}$ be the $(J + J') \times 1$ vector of the mean and dispersion parameters. The definition of the likelihood-like function $L(T; \boldsymbol{\theta})$ given for standard GEE holds using the more general parameter vector $\boldsymbol{\theta}$. Differentiate $\ell(T; \boldsymbol{\theta}) = \log(L(T; \boldsymbol{\theta}))$ with respect to the vector $\boldsymbol{\beta}'$ of dispersion coefficient parameters while holding the correlation parameter ρ fixed in the current parameter vector $\boldsymbol{\beta}'$ to provide the J' estimating equations

$g(\boldsymbol{\beta}') = \partial'\ell(T;\boldsymbol{\theta})/\partial'\boldsymbol{\beta}' = \mathbf{0}$ where the notation $\partial'\ell(T;\boldsymbol{\theta})/\partial'\boldsymbol{\beta}$ is used to indicate that this is not the full partial derivative vector for $\ell(T;\boldsymbol{\theta})$ in $\boldsymbol{\beta}'$ due to not accounting for the effect of $\boldsymbol{\beta}'$ on ρ. Now, combine these with the J standard GEE equations $g(\boldsymbol{\beta}) = \mathbf{0}$ to solve for joint estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$. Then, iteratively solve for

$$g(\boldsymbol{\theta}) = \left(g(\boldsymbol{\beta})^{\mathrm{T}} \ g(\boldsymbol{\beta}')^{\mathrm{T}}\right)^{\mathrm{T}} = \mathbf{0}$$

with $g(\boldsymbol{\theta})$ in the role of the gradient vector and the $(J+J') \times (J+J')$ matrix $H(\boldsymbol{\theta})$ in the role of the Hessian matrix. $H(\boldsymbol{\theta})$ has four component submatrices: the $J \times J$ matrix $H(\boldsymbol{\beta})$ for the mean coefficients as defined for standard GEE, the $J' \times J'$ matrix $H(\boldsymbol{\beta}') = \partial'g(\boldsymbol{\beta}')/\partial'\boldsymbol{\beta}'$ for the J' dispersion coefficients, the $J \times J'$ matrix $H(\boldsymbol{\beta},\boldsymbol{\beta}') = \partial'g(\boldsymbol{\beta})/\partial'\boldsymbol{\beta}'$, and its transpose $H(\boldsymbol{\beta}',\boldsymbol{\beta}) = H(\boldsymbol{\beta},\boldsymbol{\beta}')^{\mathrm{T}}$.

Note that

$$\log_e(|\boldsymbol{\Sigma}|) = \log_e(|\mathbf{R}(\rho)|) + \sum_{i=1}^{N} \log_e(\varphi_{t(i)}) + \sum_{i=1}^{N} \log_e\left(V\left(\mu_{t(i)}\right)\right),$$

$$\varphi_{t(i)} = \exp\left(x_{t(i)}'^{\mathrm{T}} \cdot \boldsymbol{\beta}'\right)$$

and

$$e^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot e = stde^{\mathrm{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot stde$$

where $stde$ is the $N \times 1$ vector with entries $stde_{t(i)} = e_{t(i)}/\sigma_{t(i)}$ for $t(i) \in T$. Consequently, $g(\boldsymbol{\beta}')$ has entries

$$g_j(\boldsymbol{\beta}') = stdex_j'^{\mathrm{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot stde - \sum_{i=1}^{N} x_{t(i),j}'/2$$

for $1 \le j \le J'$ where $stdex_j'$ is the $N \times 1$ vector with entries

$$stdex_{t(i),j}' = x_{t(i),j}' \cdot stde_{t(i)}/2$$

for $t(i) \in T$. $H(\boldsymbol{\beta}')$ has entries

$$\mathbf{H}_{j,j'}(\boldsymbol{\beta}') = -stdexx_{j,j'}''^{\mathrm{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot stde - stdex_j'^{\mathrm{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot stdex_{j'}'$$

for $1 \le j, j' \le J'$ where $stdexx_{j,j'}''$ is the $N \times 1$ vector with entries

$$stdexx_{t(i)j,j'}'' = x_{t(i),j}' \cdot x_{t(i),j'}' \cdot stde_{t(i)}/4$$

for $t(i) \in T$. $H(\boldsymbol{\beta},\boldsymbol{\beta}')$ has columns

$$\mathbf{H}_j(\boldsymbol{\beta},\boldsymbol{\beta}') = -\mathbf{D}^{\mathrm{T}} \cdot \mathrm{Diag}\left(\boldsymbol{\sigma}invx_j'\right) \cdot \mathbf{R}^{-1}(\rho) \cdot stde$$
$$- \mathbf{D}^{\mathrm{T}} \cdot \mathrm{Diag}(1/\boldsymbol{\sigma}) \cdot \mathbf{R}^{-1}(\rho) \cdot stdex_j'$$

where $\boldsymbol{\sigma}invx_j'$ is the $N \times 1$ vector with entries $\sigma invx_{t(i),j}' = x_{t(i),j}'/\left(2 \cdot \sigma_{t(i)}\right)$ for $t(i) \in T$ and $1 \le j \le J'$. If offsets are included, they are carried along in equations without any effect on derivatives.

4.2. Parameter Estimation

Given a value for the vector $\boldsymbol{\theta}$ of all coefficient parameters, an estimate of the correlation parameter ρ can be based on the associated standardized residuals

$stde_{t(i)}$. Calculate correlation estimates for the IND, EXCH, and AR1 correlation structures using the same formulas as before but computed with these more general standardized residuals. Iteratively solve $g(\theta) = 0$ as follows. Given the current value θ_u for θ, the next value is given by $\theta_{u+1} = \theta_u - \mathbf{H}^{-1}(\theta_u) \cdot g(\theta_u)$, thereby adapting Newton's method with $g(\theta)$ in the role of the gradient vector and $\mathbf{H}(\theta)$ in the role of the Hessian matrix. The solution to the estimating equations for observations indexed by T is denoted as $\theta(T) = \left(\boldsymbol{\beta}(T)^{\mathrm{T}} \boldsymbol{\beta}'(T)^{\mathrm{T}} \right)^{\mathrm{T}}$ with associated correlation estimate $\rho(T) = \rho(\theta(T))$.

5. Extended Linear Mixed Modeling

5.1. Formulation

GEE modeling can be further extended to handle full parameter estimation through maximizing the likelihood-like function. Let $\theta = \left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\beta}'^{\mathrm{T}} \rho \right)^{\mathrm{T}}$ be the $(J + J' + 1) \times 1$ vector of the mean, dispersion, and correlation parameters. The definition of the likelihood-like function $L(T; \theta)$ given for standard GEE holds using this more general parameter vector θ. The likelihood-like function $L(T; \theta)$ is maximized in the coefficient parameter vector θ by solving the estimating equations

$$g(\theta) = \partial \ell(T; \theta) / \partial \theta = \mathbf{0}$$

where $\partial \ell(T; \theta) / \partial \theta$ is the vector of standard partial derivatives of $\ell(T; \theta)$. The associated matrix $\mathbf{H}(\theta) = \partial g(\theta) / \partial \theta$. In this case, $g(\theta)$ is a true gradient vector and $\mathbf{H}(\theta)$ a true Hessian matrix. This approach is extended linear mixed modeling in the sense that if the entries of y were continuous variables treated as normally distributed with $V(\mu) = 1$, then it would be exactly linear mixed modeling. Formulations given in what follows are adapted from those of [8].

The gradient vector $g(\theta) = \left(g(\boldsymbol{\beta})^{\mathrm{T}} g(\boldsymbol{\beta}')^{\mathrm{T}} g(\rho) \right)^{\mathrm{T}}$. The gradient sub-vector $g(\boldsymbol{\beta}') = \partial \ell(T; \theta) / \partial \boldsymbol{\beta}'$ has the same formulation as for extended GEE modeling, only now its entries are standard partial derivatives. The gradient subvector $g(\boldsymbol{\beta}) = \partial \ell(T; \theta) / \partial \boldsymbol{\beta}$ has entries

$$g_j(\boldsymbol{\beta}) = \boldsymbol{stdex}_j^{\mathrm{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot \boldsymbol{stde} - \sum_{i=1}^{N} x_{t(i),j} / 2$$

where $\boldsymbol{stdex}_j$ is the $N \times 1$ vector with entries

$$stdex_{t(i),j} = x_{t(i),j} \left(y_{t(i)} + \mu_{t(i)} \right) / \left(2 \cdot \sigma_{t(i)} \right)$$

for $t(i) \in T$ and $1 \le j \le J$. The partial derivative $g(\rho) = \partial \ell(T; \theta) / \partial \rho$ satisfies

$$g(\rho) = -\boldsymbol{stde}^{\mathrm{T}} \cdot \partial \mathbf{R}^{-1}(\rho) / \partial \rho \cdot \boldsymbol{stde} / 2 - \partial \left(\log \left(|\mathbf{R}(\rho)| \right) \right) / \partial \rho / 2$$

where

$$\partial \left(\log \left(|\mathbf{R}(\rho)| \right) \right) / \partial \rho = tr \left(\mathbf{R}^{-1}(\rho) \cdot \partial \mathbf{R}(\rho) / \partial \rho \right),$$

tr denotes the trace function, and

$$\partial \mathbf{R}^{-1}(\rho)/\partial \rho = -\mathbf{R}^{-1}(\rho) \cdot \partial \mathbf{R}(\rho)/\partial \rho \cdot \mathbf{R}^{-1}(\rho).$$

For IND correlations, $\partial \mathbf{R}(\rho)/\partial \rho = \mathbf{0}$. For EXCH correlations, $\partial \mathbf{R}(\rho)/\partial \rho$ is the $N \times N$ matrix with diagonal entries all equal to 0 and off-diagonal entries all equal to 1. For AR1 correlations, $\partial \mathbf{R}(\rho)/\partial \rho$ is the $N \times N$ matrix with diagonal entries all equal to 0 and off-diagonal entries equaling

$$\left| t(i) - t(i') \right| \cdot \rho_{\text{AR1}}^{|t(i)-t(i')|-1}$$

in the ith row and i'^{th} column for $1 \le i \ne i' \le N$.

$H(\boldsymbol{\beta})$ has nine component submatrices: the $J \times J$ matrix $H(\boldsymbol{\beta}) = \partial g(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ for the mean parameters, the $J' \times J'$ matrix $H(\boldsymbol{\beta}') = \partial g(\boldsymbol{\beta}')/\partial \boldsymbol{\beta}'$ for the dispersion parameters computed as for extended GEE modeling, the second partial derivative $H(\rho) = \partial g(\rho)/\partial \rho$ for the correlation parameter, the $J \times J'$ matrix $H(\boldsymbol{\beta}, \boldsymbol{\beta}') = \partial g(\boldsymbol{\beta})/\partial \boldsymbol{\beta}'$, and its transpose $H(\boldsymbol{\beta}', \boldsymbol{\beta}) = H(\boldsymbol{\beta}, \boldsymbol{\beta}')^{\text{T}}$, the $J \times 1$ vector $H(\boldsymbol{\beta}, \rho) = \partial g(\boldsymbol{\beta})/\partial \rho$ and its transpose $H(\rho, \boldsymbol{\beta}) = H(\boldsymbol{\beta}, \rho)^{\text{T}}$, and the $J' \times 1$ vector $H(\boldsymbol{\beta}', \rho) = \partial g(\boldsymbol{\beta}')/\partial \rho$ and its transpose $H(\rho, \boldsymbol{\beta}') = H(\boldsymbol{\beta}', \rho)^{\text{T}}$. $H(\boldsymbol{\beta})$ has entries

$$\mathbf{H}_{j,j'}(\boldsymbol{\beta}) = -\boldsymbol{stdexx}_{j,j'}^{\text{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot \boldsymbol{stde} - \boldsymbol{stdex}_{j}^{\text{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot \boldsymbol{stdex}_{j'}$$

for $1 \le j, j' \le J'$ where $\boldsymbol{stdexx}_{j,j'}$ is the $N \times 1$ vector with entries

$$stdexx_{t(i)j,j'} = x_{t(i),j} \cdot x_{t(i),j'} \cdot stde_{t(i)}/4$$

for $t(i) \in T$. The second partial derivative $H(\rho)$ satisfies

$$H(\rho) = -\boldsymbol{stde}^{\text{T}} \cdot \partial^2 \mathbf{R}^{-1}(\rho)/\partial \rho^2 \cdot \boldsymbol{stde}/2 - \partial^2 \left(\log\left(\left| \mathbf{R}(\rho) \right| \right) \right)/\partial \rho^2 /2$$

where

$$\partial^2 \left(\log\left(\left| \mathbf{R}(\rho) \right| \right) \right)/\partial \rho^2 = -tr\left(\mathbf{R}^{-1}(\rho) \cdot \partial \mathbf{R}(\rho)/\partial \rho \cdot \mathbf{R}^{-1}(\rho) \cdot \partial \mathbf{R}(\rho)/\partial \rho \right)$$
$$+ tr\left(\mathbf{R}(\rho)^{-1} \cdot \partial^2 \mathbf{R}(\rho)/\partial \rho^2 \right).$$

For IND and EXCH correlations, $\partial^2 \mathbf{R}(\rho)/\partial \rho^2 = 0$. For AR1 correlations, $\partial^2 \mathbf{R}(\rho)/\partial \rho^2$ is the $N \times N$ matrix with diagonal entries all equal to 0 and off-diagonal entries equaling

$$\left| t(i) - t(i') \right| \cdot \left(\left| t(i) - t(i') \right| - 1 \right) \cdot \rho_{\text{AR1}}^{|t(i)-t(i')|-2}$$

in the ith row and i'^{th} column for $1 \le i \ne i' \le N$. $H(\boldsymbol{\beta}', \boldsymbol{\beta})$ has entries

$$\mathbf{H}_{j,j'}(\boldsymbol{\beta}', \boldsymbol{\beta}) = -\boldsymbol{stdexx}_{j,j'}'^{\text{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot \boldsymbol{stde} - \boldsymbol{stdex}_{j}^{\text{T}} \cdot \mathbf{R}^{-1}(\rho) \cdot \boldsymbol{stdex}_{j'}'$$

for $1 \le j, j' \le J'$ where $\boldsymbol{stdexx}_{j,j'}'$ is the $N \times 1$ vector with entries

$$stdexx_{t(i)j,j'}' = x_{t(i),j'}' \cdot stdex_{t(i),j}/2$$

for $t(i) \in T$, $1 \le j \le J$, and $1 \le j' \le J'$. $H(\boldsymbol{\beta}, \rho)$ has entries

$$\mathbf{H}_{j}(\boldsymbol{\beta}, \rho) = \boldsymbol{stdex}_{j}^{\text{T}} \cdot \partial \mathbf{R}(\rho)^{-1}/\partial \rho \cdot \boldsymbol{stde}$$

for $1 \le j \le J$. $H(\boldsymbol{\beta}', \rho)$ has entries

$$\mathbf{H}_{j}(\boldsymbol{\beta}', \rho) = \boldsymbol{stdex}_{j}'^{\text{T}} \cdot \partial \mathbf{R}^{-1}(\rho)/\partial \rho \cdot \boldsymbol{stde}$$

for $1 \le j \le J'$.

5.2. Parameter Estimation

The parameter vector $\boldsymbol{\theta}$ is estimated by iteratively solving $\boldsymbol{g}(\boldsymbol{\theta}) = 0$ as follows. Given the current value $\boldsymbol{\theta}_u$ for $\boldsymbol{\theta}$, the next value is given by

$$\boldsymbol{\theta}_{u+1} = \boldsymbol{\theta}_u - \mathbf{H}^{-1}(\boldsymbol{\theta}_u) \cdot \boldsymbol{g}(\boldsymbol{\theta}_u),$$

thereby using Newton's method with gradient vector $\boldsymbol{g}(\boldsymbol{\theta})$ and Hessian matrix $\mathbf{H}(\boldsymbol{\theta})$. The estimation process can be stopped early if $\ell(T; \boldsymbol{\theta}_{u+1})$ does not increase by much compared to $\ell(T; \boldsymbol{\theta}_u)$. The solution to the estimating equations for observations indexed by T is denoted as $\boldsymbol{\theta}(T) = \left(\boldsymbol{\beta}(T)^{\mathrm{T}} \, \boldsymbol{\beta}'(T)^{\mathrm{T}} \, \rho(T) \right)^{\mathrm{T}}$.

The covariance matrix for the parameter estimate vector $\boldsymbol{\theta}(T)$ can be computed as $-\mathbf{H}^{-1}(\boldsymbol{\theta}(T))$ and the variances corresponding to its diagonal entries can be used to compute z tests of zero individual model parameters. These are useful for fixed models of theoretical importance. On the other hand, tests for parameters of adaptively generated models (as described in Section 6) are usually significant as a consequence of the model selection process, and so results for these tests are not reported for models generated in the example analyses.

## 6. Modeling Possibly Nonlinear Means and Dispersion over Time

Knafl and Ding [3] provide a detailed formulation for adaptively searching through alternative regression models for means and dispersions in a variety of contexts using adaptive fractional polynomial models [9]. A brief overview is provided here. These methods are used in the example analyses of individual cancer patient count/rate data presented later. Model selection proceeds through two phases. The expansion phase first grows the model adding in alternative power transforms of predictors for means and dispersions. The contraction phase then reduces the model to a parsimonious set of power transforms by removing transforms from the current model one at a time and adjusting the powers of the remaining transforms. Alternative models are evaluated using LCV scores. The modeling process is controlled by tolerance parameters indicating how much of a reduction in the LCV score can be tolerated at given stages of the process. Knafl and Ding [3] also provide a wide variety of example analyses demonstrating the usefulness of these adaptive regression methods. A description of these methods in the standard Poisson regression context is provided in [10].

A SAS® (SAS Institute, Inc., Cary, NC) macro has been developed for generating adaptive analyses including the reported example analyses. This macro as well as data and code used to generate the results of the example analyses are available from the first author.

7. Example Analyses

7.1. Pain Flare Counts per Day

Figure 1 displays pain flare counts for Cancer Patient 1 over a period of 34 days.
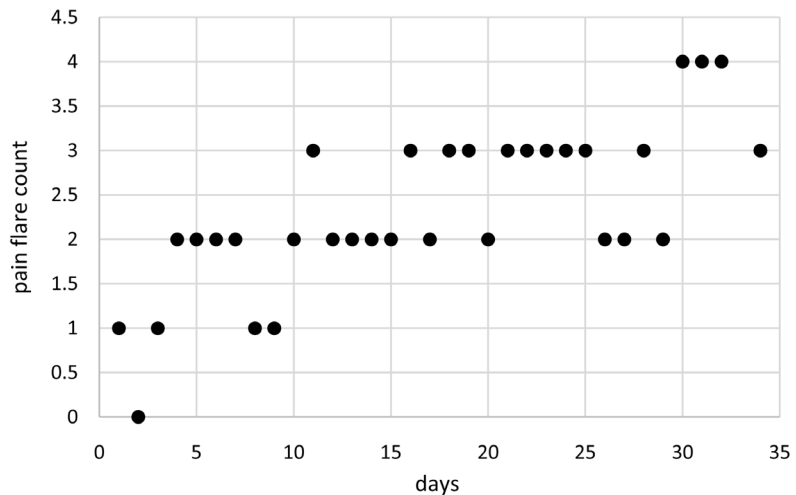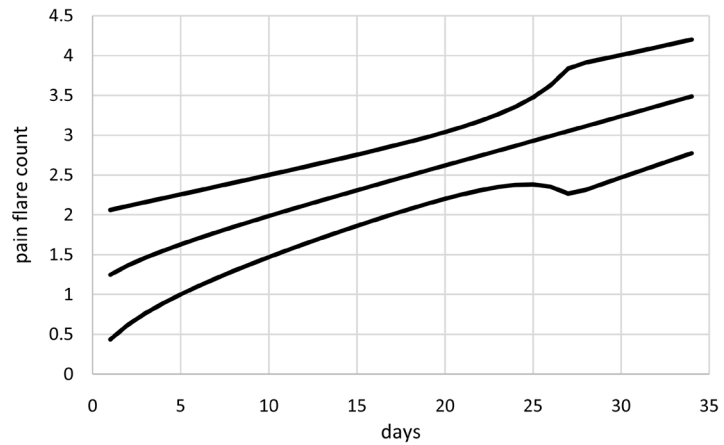
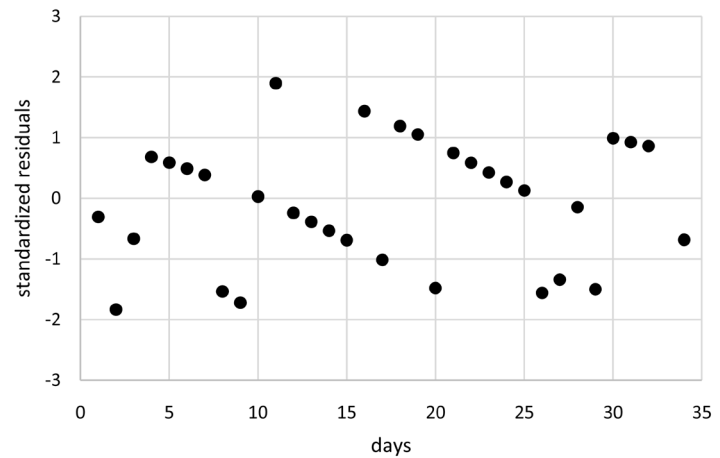**Figure 1.** Pain flare counts over time for cancer patient 1.

Pain flares range from 0 to 4 per day and tend to increase over time. Data are available for $N = 33$ days with a missing value for one day (day 33). These data were collected using Ecological Momentary Assessment (EMA) [11] as implemented in the mEMA app [12].

Table 1 contains results for adaptive models for means and dispersions of pain flare counts over time using the two modeling approaches extended GEE modeling and extended linear mixed modeling and the three correlation structures IND, AR1, and EXCH. Power transforms reported in Table 1 were generated by adaptively searching through alternative power transforms using the methods described in Section 6. LCV scores are based on $k = 5$ folds with fold sizes ranging from 2 to 8 measurements and no empty folds. For extended GEE modeling, IND correlations generate the best LCV score 0.38018 over the three correlation structures. For extended linear mixed modeling, IND correlations also generate the best LCV score 0.40622. These results suggest that a Poisson process assumption is reasonable for these pain flare counts.

Extended linear mixed modeling generates better LCV scores than extended GEE for all three correlation structures. Moreover, computation times are much shorter ranging from 0.4 to 1.2 minutes compared to 13.9 to 35.5 minutes. These results suggest that extended linear mixed modeling is preferable for modeling these pain flare counts because it generates better LCV scores in less time. Consequently, only extended linear mixed modeling using IND correlations is considered further for these data, generating the model with means based on $t(i)^{0.49}$ without an intercept and dispersions based on $t(i)^{8.37}$ and $t(i)^{0.5}$ without an intercept. Figure 2 displays estimates of mean pain flare counts over time along with unit error bands over time (*i.e.*, the mean ±1 extended standard deviation at each time) to account for variability about the means. Mean pain flare counts increase over time, somewhat close to linearly in time. Variability in pain flare counts is smaller in the middle of the period, somewhat larger at the start of the period and even larger at the end of the period. Figure 3 displays

Figure 2. Mean pain flare counts (middle curve) with unit error bounds for cancer patient 1.

Figure 3. Standardized residuals for the model of **Figure 2**.

Table 1. Adaptive models for means and dispersions of pain flare counts over time for alternative modeling approaches and correlation structures.

Modeling Approach	Correlation		Power Transforms[a]		5-fold LCV Score	Time[b]
	Structure	Estimate	Means	Dispersions		
extended GEE	IND	0	$t(i)^{0.54}$	$t(i)^{0.12}$	038018	13.9
	EXCH	0.001	$t(i)^{0.869}$	$t(i)^{0.29}$	0.34712	35.5
	AR1	0.10	$t(i)^{0.59}$	$t(i)^{0.1}$	0.37404	17.6
extended LMM	IND	0	$t(i)^{0.49}$	$t(i)^{8.37}$, $t(i)^{0.5}$	0.40622	0.4
	EXCH	0.42	$t(i)^{0.511}$	$t(i)^{0.2}$	0.36590	1.2
	AR1	0.20	$t(i)^{0.4}$	1, $t(i)^{1.01}$, $t(i)^{1.01}$	0.37693	0.4

AR1—autoregressive of order 1; EXCH—exchangeable; GEE—generalized estimating equations; IND—independent; LCV—likelihood-like cross-validation; LMM—linear mixed modeling. a. The ith time value is denoted as $t(i)$. A 1 corresponds to an intercept parameter; otherwise, the model has a zero intercept. b. Difference in minutes of clock times between the start and end of computations.

standardized residuals for this model, which range between ±2 without any extreme outliers, suggesting the model is a reasonable fit for these data.

The associated model generated using $k = 10$ folds has similar means based on $t(i)^{0.5}$ without an intercept and simpler dispersions based on $t(i)^{0.2}$ without an intercept. However, the 10-fold LCV score 0.38107 is smaller, suggesting that $k = 5$ is a better choice for these data. Moreover, there is one empty fold, suggesting that the choice of $k = 10$ folds is too large for these data with only $N = 33$ measurements. The associated model generated with $k = 5$ folds and assuming constant dispersions has a similar model for the means based on $t(i)^{0.53}$ without an intercept but a smaller LCV score 0.37031, suggesting that the dispersions for these data are reasonably treated as nonconstant over time.

7.2. As Needed Pain Medications Taken Counts per Day

Figure 4 displays as needed pain medications taken counts for Cancer Patient 2 over a period of 100 days. As needed pain medications taken counts range from 0 to 4 per day and tend to decrease over time. Data are available for $N = 92$ days with a missing value for eight other days (days 4, 14, 52, 56, 74, 81, 85, and 89). These data were also collected using the mEMA app.

Table 2 contains results for adaptive models for means and dispersions of as needed pain medications taken counts over time using the two modeling approaches extended GEE modeling and extended linear mixed modeling and the three correlation structures IND, AR1, and EXCH. LCV scores are based on $k = 5$ folds with fold sizes ranging from 13 to 21 measurements with no empty folds. For extended GEE modeling, AR1 correlations generate the best LCV score 0.41497 over the three correlation structures. For extended linear mixed modeling, AR1 correlations generate the best LCV score 0.40509. These results indicate that a Poisson process assumption may not be appropriate for these as needed pain medications taken counts.
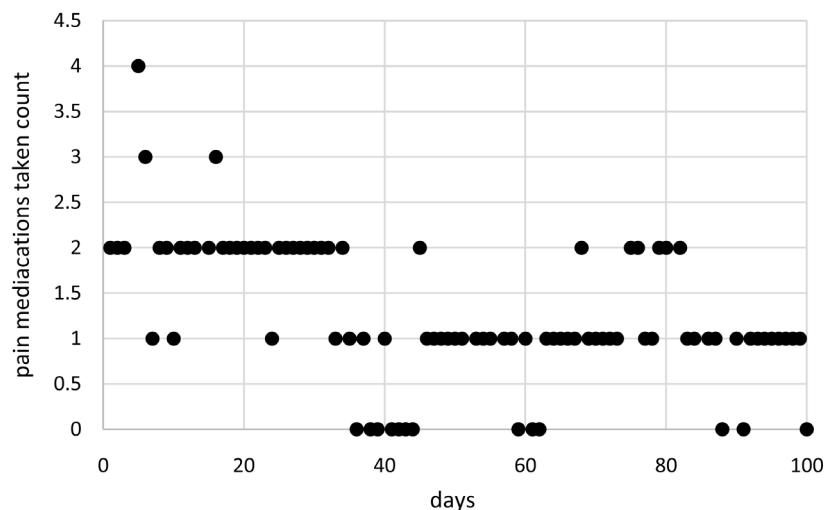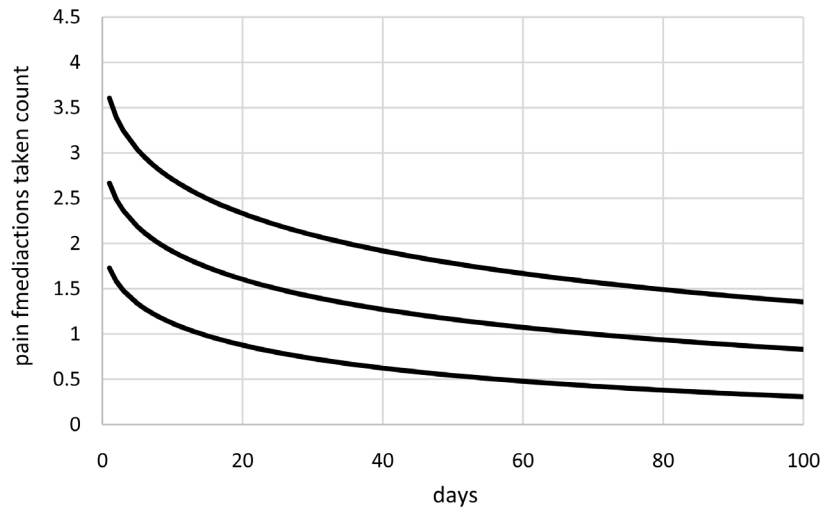

Figure 4. As needed pain medications taken counts over time for cancer patient 2.

Table 2. Adaptive models for means and dispersions of as needed pain medications taken counts over time for alternative modeling approaches and correlation structures.

Modeling Approach	Correlation		Power Transforms[a]		5-fold LCV Score	Time[b]
	Structure	Estimate	Means	Dispersions		
extended GEE	IND	0	$1, \ t(i)^{0.5}$	1	037030	84.5
	EXCH	−0.01	$1, \ t(i)^{0.5}$	1	0.36340	202.3
	AR1	0.57	$t(i)^{-0.12}, \ t(i)^3$	$1, \ t(i)^{-1.5}$	0.41497	222.7
extended LMM	IND	0	$t(i)^{0.3}, \ t(i)^{0.1}$	1	0.36845	0.5
	EXCH	−0.01	$1, \ t(i)^{0.4}$	1	0.37379	1.7
	AR1	0.45	$1, \ t(i)^{0.4}$	1	0.40509	0.8

AR1—autoregressive of order 1; EXCH—exchangeable; GEE—generalized estimating equations; IND—independent; LCV—likelihood-like cross-validation; LMM—linear mixed modeling. a. The ith time value is denoted as $t(i)$. A 1 corresponds to an intercept parameter; otherwise, the model has a zero intercept. b. Difference in minutes of clock times between the start and end of computations.

Extended GEE modeling generates a better LCV score than extended linear mixed modeling for the IND correlation structure, but the scores for these two approaches are not too different. Extended linear mixed modeling generates a better LCV score than extended GEE modeling for the EXCH correlation structure. Extended GEE modeling generates a better LCV score than extended linear mixed modeling for the AR1 correlation structure. Although this is the best overall LCV score, the associated model for extended linear mixed modeling is more parsimonious with an intercept and one time transform for the means compared to two time transforms and constant dispersions compared to dispersions based on and intercept and one time transform. Moreover, computation times are substantially shorter for extended linear mixed modeling ranging from 0.5 to 1.7 minutes compared to 84.5 to 222.7 minutes or 1.4 to 3.7 hours. These results suggest that extended linear mixed modeling is preferable for modeling these as needed pain medications taken counts because it generates competitive or better scores or more parsimonious models in substantially less time. Consequently, only extended linear mixed modeling using AR1 correlations are considered further for these data, generating the model with means based on $t(i)^{0.4}$ with an intercept, constant dispersions based on an intercept, and estimated autocorrelation $\rho_{AR1} = 0.45$. **Figure 5** displays estimates of mean as needed pain medications taken counts over time along with unit error bands over time (*i.e.*, the mean ±1 extended standard deviation at each time) to account for variability about the means. Mean as needed pain medications taken counts decrease non-linearly over time. Variability in as needed pain medications taken counts is close to constant over time. **Figure 6** displays standardized residuals for this model, which range well within ±3 without any extreme outliers, suggesting the model provides a reasonable fit to these data.

Figure 5. Mean as needed pain medications taken counts (middle curve) with unit error bounds for cancer patient 2.
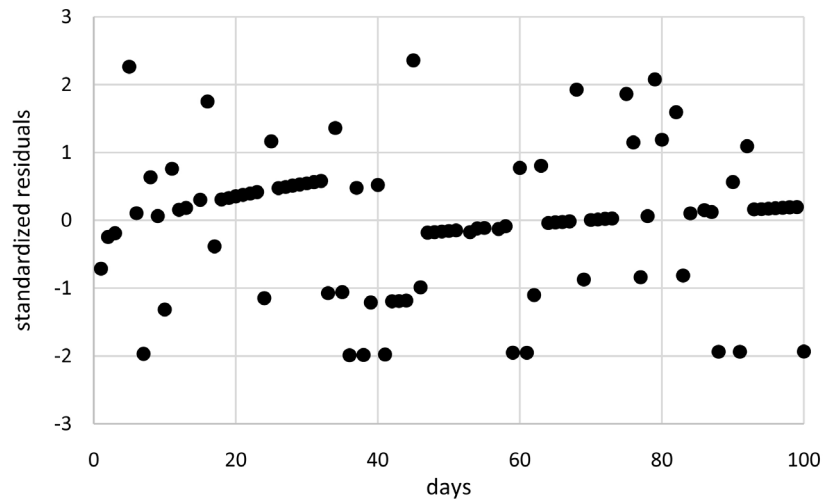

Figure 6. Standardized residuals for the model of **Figure 5**.

The associated model generated using $k = 10$ folds is about the same with means based on $t(i)^{0.5}$ with an intercept, constant dispersions, and estimated correlation $\rho_{AR1} = 0.45$. The 10-fold LCV score 0.40958 is larger, suggesting that $k = 10$ is a better choice for these data. There are no empty folds. The associated model generated using $k = 15$ folds is similar with means based on $t(i)^{0.5}$ with an intercept, dispersions based on $t(i)^{0.07}$ without an intercept, and estimated correlation $\rho_{AR1} = 0.46$. The 15-fold LCV score 0.40353 is smaller, suggesting that $k = 10$ is a better choice for these data. There are no empty folds. The associated model generated with $k = 15$ folds assuming constant dispersions has means based on $t(i)^{0.5}$ with an intercept and close 15-fold LCV score 0.40318. Consequently, models generated by 5, 10, and 15 folds using extended linear mixed modeling are not too different, suggesting that the results are reasonably robust to the choice of the number of folds.

7.3. Around the Clock Pain Medications Taken Rates per Day per Dose

Adherence data for around the clock pain medications were collected using pill bottles equipped with Medication Event Monitoring System (MEMS) devices (AARDEX North America, Boulder, CO) that recorded the date and time of each pill bottle opening and presumably of the taking of the pain medication [13] [14]. Cancer Patient 3 was monitored for a period of 91 days. Counts of around the clock pain medications taken were computed for 30 equal-sized sub-periods of 3.03 days each, ranging from 0 to 18. Around the clock pain medications were to be taken five times a day by this patient. Methods for modeling such data assuming the special case of a Poisson process with constant dispersions are provided in [10]. Figure 7 displays around the clock pain medications taken rates per day per dose for Cancer Patient 3. The ideal rate of 1 means that the patient took around the clock pain medications at the appropriate rate over the associated time subperiod. Around the clock pain medications taken rates range from 0 to 1.19 per day per dose and tend to decrease over time. Data are available for $N = 30$ subperiods with none missing.

Table 3 contains results for adaptive models for means and dispersions of around the clock pain medications taken rates over time using the two modeling approaches extended GEE modeling and extended linear mixed modeling and the three correlation structures IND, AR1, and EXCH. LCV scores are based on $k = 5$ folds with fold sizes ranging from 2 to 8 measurements with no empty folds. For extended GEE modeling, EXCH correlations generate the best LCV score 0.051583 over the three correlation structures. For extended linear mixed modeling, AR1 correlations generate the best LCV score 0.053856, which is also the best overall LCV score for Table 3 models. For both modeling approaches, the LCV score for IND correlations is quite a bit smaller than the best LCV score over the three correlation structures. These results indicate that a Poisson process assumption may not be appropriate for these around the clock pain medications taken rates.
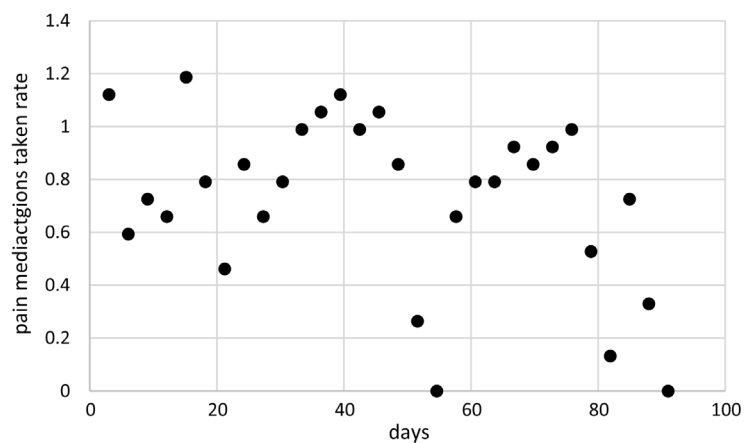
Figure 7. Around the clock pain medications taken rates per day per dose over time for cancer patient 3.

Table 3. Adaptive models for means and dispersions of around the clock pain medications taken rates per day per dose over time for alternative modeling approaches and correlation structures.

| Modeling Approach	Correlation		Power Transforms[a]		5-fold LCV Score	Time[b]
	Structure	Estimate	Means	Dispersions		
extended GEE	IND	0	$t(i)^{0.7}$	1	0.046556	5.1
	EXCH	−0.03	$1,\ t(i)^{5}$	1	0.051583	11.9
	AR1	0.58	$t(i)^{1.1}$	1	0.048525	6.9
extended LMM	IND	0	$t(i)^{0.8}$	1	0.046837	0.2
	EXCH	−0.03	$t(i)^{0.9}$	1	0.045251	0.7
	AR1	0.75	$t(i)^{1.1}$	$1,\ t(i)^{-1.5}$	0.053856	0.2

AR1—autoregressive of order 1; EXCH—exchangeable; GEE—generalized estimating equations; IND—independent; LCV—likelihood-like cross-validation; LMM—linear mixed modeling. a. The ith time value is denoted as $t(i)$. A 1 corresponds to an intercept parameter; otherwise, the model has a zero intercept. b. Difference in minutes of clock times between the start and end of computations.

Extended linear mixed modeling generates better LCV scores than extended GEE for the IND and AR1 correlation structures. Its LCV score is smaller for the EXCH correlation structure, but its model is more parsimonious based on one time transform for the means with constant dispersions compared to one time transform plus an intercept for the means with constant dispersions. Furthermore, computation times are much shorter for extended linear mixed modeling ranging from 0.2 to 0.7 minutes compared to 5.1 to 11.9 minutes. These results suggest that extended linear mixed modeling is preferable for modeling these around the clock pain medications taken rates because it generates the best LCV score in less time. Consequently, only extended linear mixed modeling using AR1 correlations are considered further for these data, generating the model with means based on $t(i)^{1.1}$ without an intercept, dispersions based on $t(i)^{6.1}$ with an intercept, and estimated autocorrelation $\rho_{AR1} = 0.75$. Figure 8 displays estimates of mean around the clock pain medications taken rates over time along with unit error bands over time (*i.e.*, the mean ±1 extended standard deviation at each time) to account for variability about the means. Mean around the clock pain medications taken counts decrease close to linearly over time. Variability in around the clock pain medications taken rates is larger at the end of the period. Figure 9 displays standardized residuals for this model, which range well within ±3 without any extreme outliers, suggesting the model provides a reasonable fit to these data.

The associated model generated using $k = 10$ folds is somewhat similar with means based on $t(i)^{0.4}$ without an intercept, constant dispersions based on an intercept, and estimated autocorrelation $\rho_{AR1} = 0.76$. However, the 10-fold LCV score 0.052023 is smaller, suggesting that $k = 5$ is a better choice for these data. Moreover, there is one empty fold, suggesting that the choice of $k = 10$
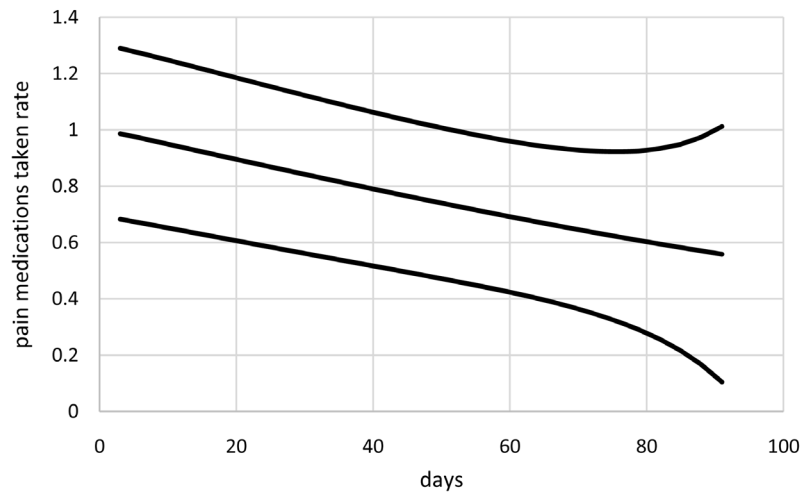
Figure 8. Mean around the clock pain medications taken rates per day per dose (middle curve) with unit error bounds for cancer patient 3.
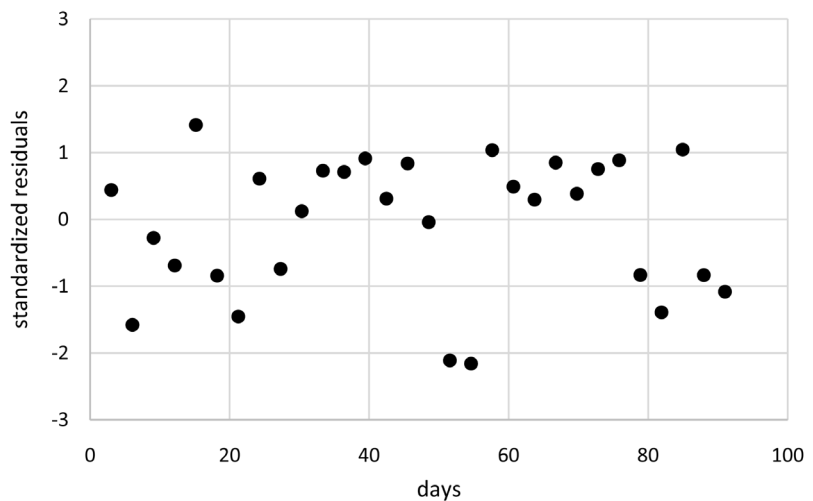
Figure 9. Standardized residuals for the model of **Figure 8**.

folds is too large for these data with only $N = 30$ measurements. The associated model generated with $k = 5$ folds and assuming constant dispersions has a model for the means based on based on $t(i)^{1.01}$ without an intercept, an autocorrelation estimate of $\rho_{AR1} = 0.75$, and a smaller LCV score 0.0.050386, suggesting that the dispersions for these data are reasonably treated as nonconstant over time.

8. Discussion

8.1. Summary

Methods are formulated for modeling individual patient count/rate data over time allowing for nonlinear trajectories for means, time-varying dispersions, and temporal correlation. Three correlation structures are considered including IND, EXCH, and spatial AR1 correlations. Two extensions of standard GEE modeling

are considered. Extended GEE modeling augments standard GEE mean parameter estimating equations with dispersion parameter estimating equations while using the GEE approach for correlation parameter estimation. Extended linear mixed modeling estimates all model parameters using estimating equations for mean, dispersion, and correlation parameters. These new estimating equations are determined by partial derivatives of a likelihood-like function based on the multivariate normal density. This likelihood-like function is also used to define a likelihood-like cross-validation (LCV) score for evaluating models. LCV scores are used to control adaptive regression modeling of possibly nonlinear means and dispersions over time. It is also possible to generate penalized likelihood-like criteria for model selection generalizing standard penalized likelihood criteria [15] such as the commonly used Akaike information criterion (AIC) and Bayesian information criterion (BIC). Pan [16] has formulated a penalized model selection criterion related to the AIC called the quasi-likelihood information criterion (QIC) for GEE model selection, but the QIC score does not fully account for the correlation structure. Model selection criteria based on the likelihood-like function fully account for the correlation structure.

Example analyses using these methods are provided using three types of count/rate data for individual cancer patients including cancer pain flares per day, as needed cancer pain medications taken per day, and around the clock cancer pain medications taken per day per dose. Extended linear mixed modeling generates models with either better LCV scores or more parsimonious models than extended GEE modeling. Moreover, times to compute models are substantially smaller for extended linear mixed modeling than for extended GEE modeling. Time differences can be extreme for even moderate samples sizes, for example, analyses for the second example data set with 92 observations required at most 1.7 minutes for extended linear mixed modeling compared to up to 3.7 hours for extended GEE. These results indicate that extended linear mixed modeling is preferable for modeling individual patient count/rate data over time. This is likely to hold in more general modeling situations with other types of data and for combined data for multiple patients.

8.2. Alternative Approaches

The formulation provided here assumes that separate modeling of each patient's longitudinal data is preferable to modeling the combined data for all patients. Separate modeling is a person-centered approach to modeling longitudinal data as opposed to a variable-centered approach using the combined data [17] [18]. This is only feasible when there are substantial numbers of time measurements for each patient. Modeling the combined data for all patients typically involves the assumption that means and dispersions for all patients are reasonably treated as having the same functional form. Knafl *et al.* [10] provide an example where this is not an appropriate assumption for a specific set of data on medication taken rates per day for HIV patients on antiretroviral medications. In any case,

the methods considered here generalize to handle combined data for multiple patients, not only count/rate longitudinal data but also continuous and dichotomous longitudinal data, and not just data for cancer patients.

Multilevel (or hierarchical linear) modeling [19] [20] could alternatively be used to provide for individual patient differences, but that usually accounts for nonlinearity using polynomial models, often simple quadratic models. Polynomial models can be too simplistic for addressing general nonlinearity. Knafl and Ding [3] provide an example for independent data where the polynomial model generating the best LCV score for degrees 0-3 is the degree 0 constant model, but a nonlinear adaptive regression model generates a much better LCV score. Future research is needed to investigate general nonlinearity using adaptive regression methods applied to multilevel models as well as to random effects models [21] and to generalized linear mixed models [22].

Spatial AR1 correlations generate better models than independent and exchangeable correlations for two of the three example data sets. This suggests consideration of autoregressive and/or moving average correlations [23] of orders more than 1. As the number of time points increases, even relatively large autocorrelations can generate small correlations for larger distances apart. For example, the third example data set had an estimated autocorrelation of $\rho_{AR1} = 0.75$ with integer time measurements ranging from 1 to 30 so that the smallest correlation is $0.75^{29} = 0.0002$. The second example data set had an even smaller estimated autocorrelation of $\rho_{AR1} = 0.45$ with an even larger range of 1 to 100 integer time measurements so that the smallest correlation is $0.45^{99} = 4.7 \times 10^{-35}$. These results suggest consideration of banded correlation autoregressive structures with zero correlations for measurements further apart than some fixed amount.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Liang, K. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. https://doi.org/10.1093/biomet/73.1.13

[2] Hardin, J. and Hilbe, J. (2013) Generalized Estimating Equations. 2nd Edition, Chapman and Hall/CRC, New York. https://doi.org/10.1201/b13880

[3] Knafl, G.J. and Ding, K. (2016) Adaptive Regression for Modeling Nonlinear Relationships. Springer International Publishing, Switzerland.

[4]   McCullagh, P. and Nelder, J.A. (1999) Generalized Linear Models. 2nd Edition, Chapman & Hall/CRC, Boca Raton.

[5]   Wedderburn, R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss—Newton Method. *Biometrika*, **61**, 439-447. https://doi.org/10.1093/biomet/61.3.439

[6]   Chaganty, N.R. (1997) An Alternative Approach to the Analysis of Longitudinal Data via Generalized Estimating Equations. *Journal of Statistical Planning and Inference*, **63**, 39-54. https://doi.org/10.1016/S0378-3758(96)00203-0

[7]   Burman, P. (1989) A Comparative Study of Ordinary Cross-Validation, ν-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika*, **76**, 503-514. https://doi.org/10.1093/biomet/76.3.503

[8]   Wolfinger, R., Tobias, R. and Sall, J. (1994) Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models. *Siam Journal on Scientific Computing*, **15**, 1294-1310. https://doi.org/10.1137/0915079

[9]   Knafl, G.J. (2018) Adaptive Fractional Polynomial Modeling. *Open Journal of Statistics*, **8**, 159-186. https://doi.org/10.4236/ojs.2018.81011

[10] Knafl, G.J., Fennie, K.P., Bova, C. and Williams, A.B. (2004) Electronic Monitoring Device Event Modeling on an Individual-Subject Basis Using Adaptive Poisson Regression. *Statistics in Medicine*, **23**, 783-801. https://doi.org/10.1002/sim.1624

[11] Shiffman, S., Stone, A.A. and Hufford, M.R. (2008) Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, **4**, 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[12] Ilumivi (n.d.) Software for Humanity. https://ilumivu.com/solutions/ecological-momentary-assessment-app/

[13] Meghani, S.H. and Knafl, G.J. (2016) Patterns of Analgesic Adherence Predict Health Care Utilization among Outpatients with Cancer Pain. *Patient Preference Adherence*, **10**, 81-98. https://doi.org/10.2147/PPA.S93726

[14] Meghani, S.H., Persico, A., Fudin, J. and Knafl, G.J. (2021) Gaps in the Use of Long-Acting Opioids within Intervals of Consecutive Days among Cancer Outpatients Using Electronic Pill Caps, *Pain Medicine*, **22**, 687-693. https://doi.org/10.1093/pm/pnaa273

[15] Sclove, S.L. (1987) Application of Model-Selection Criteria to Some Problems in Multivariate Analysis. *Psychometrika*, **52**, 333-343. https://doi.org/10.1007/BF02294360

[16] Pan, W. (2001) Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, **57**, 120-125. https://doi.org/10.1111/j.0006-341X.2001.00120.x

[17] Bergmann, L.R. and Trost, K. (2006) The Person-Centered Versus the Variable-Centered Approach: Are They Complementary, Opposites, or Exploring Different Worlds? *Merrill-Palmer Quarterly* (*Wayne State University Press*), **52**, 601-632. https://doi.org/10.1353/mpq.2006.0023

[18] Laursen, B. and Hoff, E. (2006) Person-Centered and Variable-Centered Approaches to Longitudinal Data. *Merrill-Palmer Quarterly* (*Wayne State University Press*), **52**, 377-389. https://doi.org/10.1353/mpq.2006.0029

[19] Bryk, S.W. and Raudenbush. A.S. (2002) Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Edition, Sage Publications, Thousand Oaks.

[20] Luke, D.A. (2004) Multilevel Modeling. Sage Publications, Thousand Oaks. https://doi.org/10.4135/9781412985147

[21] Laird, N.M. and Ware, J.H. (1982) Random-Effects Models for Longitudinal Data.

Biometrics, **38**, 963-974. https://doi.org/10.2307/2529876

[22] Stroup, W.W. (2012) Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. Chapman & Hall/CRC, Boca Raton.

[23] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008) Time Series Analysis: Forecasting and Control. 4th Edition, John Wiley & Sons, Hoboken.