Marginalized Transition Models for the Analysis of Longitudinal Count Data with Application to Two Clinical Trials

Keunbaik Lee,[1,*] **Michael J. Daniels,**[2,†] **Yongsung Joo**[3,‡] **and Ronald I. Shorr**[4,§]

[1]Assistant Professor, Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA

[2]Professor, Department of Statistics, University of Florida, Gainesville, FL

[3]Associate Professor, Department of Statistics, Dongguk University-Seoul, Seoul, South Korea.

[4]Professor, Department of Aging and Geriatrics, University of Florida and Director, Geriatric Research Education & Clinical Center (GRECC), Malcom Randall VAMC, Gainesville, FL

February 14, 2011

Summary. Data involving longitudinal counts are not uncommon. Here we propose new marginalized transition models for longitudinal count data. These models account for the correlation among response outcomes by embedding the marginal mean structure within a complete multivariate probability model with dependence modeled via a Markov structure; this allows inference on regression coefficients unconditional on the past. Fisher-

* *email:* klee4@lsuhsc.edu
† *email:* mdaniels@stat.ufl.edu
‡ *email:* yongsungjoo@dongguk.edu
§ *email:* rshorr@aging.ufl.edu

scoring and Quasi-Newton algorithms are developed for estimation. We examine bias in these models in small and large samples in the presence of missing data. The models are used to draw inference on a recent hospital-based falls prevention trial and a previously analyzed trial on epileptic seizures.

Key words: Generalized linear models; marginalized transition; Fisher-scoring; Markov structure.

1. Introduction

Longitudinal data are commonly collected in biomedical studies. Unlike cross-sectional studies, where a single outcome is measured for each subject, longitudinal studies collect observations over time from the same subject. Therefore, the repeated observations are correlated (Diggle et al., 2002). In longitudinal data analysis, serial correlation may not be of primary interest but it must be taken into account to make proper inferences. In this paper we will introduce new regression models for longitudinal count data.

For correlated count data, the standard Poisson model assumes equality between mean and variance. Zeger (1988) proposed a log linear model with random effects for a time series of counts using estimating equation approach. Li et al. (2007) proposed a first-order Markov transition model for longitudinal Poisson data to take into account possible influence of the previous counts, and a random intercept to capture baseline heterogeneity across subjects. However, even after inclusion of random effects, there still might be overdispersion. Similar to Zeger (1988), Thall and Vail (1990) proposed a mixed effects approach for modeling longitudinal count data with overdispersion. Although this mixed effects approach accommodates the overdispersion in count data, it is not able to model serial correlations such as AR(1) structures, which typically are present in longitudinal data. Jowaheer and Sutradhar (2002) developed joint generalized estimating equations for the regression parameters and the overdispersion parameter for longitudinal count data when the overdispersed repeated responses exhibit an autocorrelation structure. Booth et al. (2003) extended the negative binomial loglinear model to accommodate longitudinal overdispersed count data using random effects in the linear predictor. Jowaheer (2006) examined the efficiency of the estimates of the variance component of random effects in Poisson mixed models. Molenberghs et al. (2007) proposed a log linear model to accommodate overdispersion and clustering through normal and gamma random effects, respectively. Kaciroti et al. (2008) proposed a Bayesian transition Markov model of first order with random intercept, similar to the model by Zeger and Qaqish (1988). Alosh (2010) reviewed several approaches, including generalized estimating equation (GEE) methods, generalized linear mixed models, and the first-order integer-valued autoregressive model, for modeling longitudinal count data with dropouts. In our proposed models, we use a Markovian structure to account for the serial correlation of longitudinal outcomes. However, we focus on likelihood-based *marginalized* models.

Recently, marginalized likelihood-based models have been proposed to analyze longitudinal categorical data. The correlation among observations is explained by embedding the marginal mean structure within a complete multivariate probability model with dependence, modeled via random effects (Heagerty, 1999; Lee and Daniels, 2008; Lee et al., 2009, 2010, 2011) or a Markov structure (Heagerty, 2002; Lee and Daniels, 2007; Lee and Mercante, 2010). We focus on marginalized likelihood-based models using a Markov structure (and random effects) and extend this approach to accommodate longitudinal (overdispersed) count data using a loglinear model. There are advantages of the marginalized likelihood-based directly specified marginal models over conditional models. First, the interpretation of regression coefficients does not depend on specification of the dependence in the model unlike in conditional models. Second, they can be much less susceptible to bias resulting from random effects model mis-specification (Heagerty and Zeger, 2000; Heagerty and Kurland, 2001; Lee and Daniels, 2007; Lee et al, 2009; Lee and Mercante, 2010). Third, likelihood based inference is valid under ignorability and the missing data mechanism (mdm) need not be explicitly specified. The use of a likelihood based approach will have advantages for longitudinal data that is missing at random (MAR) and in particular, ignorable. However, semiparametric GEE approaches require explicit specification of the mdm for MAR (Robins et al., 1995). In addition, the re-weighting based approaches (based on the mdm) to handle MAR in GEEs only 'impute' missing values at the observed data points. Likelihood based approaches do not have this restriction and allow imputations outside the observed data.

The proposed models are motivated by a new prevention trial on falls (Shorr et al, 2011) and also are used to analyze a previously explored trial on epileptic seizures (Thall and Vail, 1990). The goal of the first trial was to test whether an intervention, increasing the use bed alarms, would reduce the risk of falls in hospitalized patients. A cluster randomized trial was conducted on sixteen 25-bed general medical-surgical nursing units in a single hospital to test the intervention. Eight nursing units were randomized to receive an intervention to increase alarm use and the other eight units utilized existing nursing care methods to minimize falls. The primary outcome measure was patient falls but other objectives involved finding unit-level factors (such as staffing) and patient level factors (such as medication use and case mix) associated with patient falls. A complication was that two units unexpectedly closed during the trial but excluding those units from the analysis would result in a loss of efficiency.

For both these datasets we want to allow serial correlation (via a Markov structure), but we want to assess covariate effects *unconditional* on the counts during the previous time period. This provides additional motivation for the proposed models.

The paper is arranged as follows. We introduce the marginalized transition models for longitudinal Poisson data in Section 2. In Section 3, we extend the models to accommodate longitudinal count data with overdispersion. We examine the bias of estimates and the robustness of inference on marginal mean parameters to misspecification of dependence model in simulation studies in Section 4. In Section 5, we used these models to draw inference on the falls and seizures studies. Finally, a brief summary and extensions are included in Section 6

2. Marginalized Models for Longitudinal Poisson Data

In this section, we propose several marginalized models for longitudinal poisson data, a Poisson marginalized transition model (PMTM) and a Poisson marginalized transition random effects models (PMTREM), respectively. First, we introduce the PMTM.

Let $Y_i = (Y_{i1}, \cdots, Y_{in_i})$ be a vector of longitudinal poisson responses on subject $i = 1, \cdots, N$ at times $t = t_1, \cdots, t_{n_i}$. We assume that associated exogenous but possibly time-varying covariates, $x_{it} = (x_{it1}, \cdots, x_{itr})$, are recorded for each subject at each time and that the regression model properly specifies the full covariate conditional probability such that $P(Y_{it} = y_{it}|X_{it}) = P(Y_{it} = y_{it}|X_{i1}, \cdots, X_{in_i})$.

2.1 *Poisson Marginalized Transition Model*

Let $\mu_{it}^M = E(Y_{it}|x_{it})$ and $\mu_{it}^c = E(Y_{it}|Y_{it-1})$. Then the *Poisson marginalized transition model*, PMTM, is specified using the following two regressions,

$$\text{mean model:} \qquad \log(\mu_{it}^M) = x_{it}^T\beta, \tag{1}$$

$$\text{dependence model:} \qquad \log(\mu_{it}^c) = \triangle_{it} + \gamma_{it}\{\log\max(\delta, Y_{it-1}) - \log\mu_{it-1}^M\}, \tag{2}$$

where β is the $r \times 1$ vector of regression coefficients, $\gamma_{it} = z_{it}^T\alpha$, z_{it} is a $q \times 1$ vector of subset of x_{it}, $\alpha = (\alpha_1, \cdots, \alpha_q)^T$, $0 < \delta < 1$, and $Y_{i,t-1}$ is the previous response. Here, β is a marginal mean parameter which is of the most interest in making inferences on the covariates effects. The parameter α indicates the influence of the previous counts through the logarithm of the residual, $(\log \max(\delta, Y_{it-1}) - \log \mu_{it-1}^M)$, where $\max(\delta, Y_{it-1})$ is used to ensure a positive value for logarithm (Zeger and Qaqish, 1988; Li et al. 2007).

Given β and α, we can calculate Δ_{it} from the following relationship,

$$\mu_{it}^M = \sum_{j=0} \mu_{it}^c P(Y_{it-1} = j). \tag{3}$$

A closed form expression for Δ_{it} is found in the following theorem.

Theorem I. The parameters Δ_{it} in (2) are finite and given by

$$\triangle_{it} = \log \mu_{it}^M + \gamma_{it} \log \mu_{it-1}^M + \mu_{it-1}^M - \log \left\{ \delta^{\gamma_{it}} + \sum_{j=1}^{\infty} j^{\gamma_{it}} \frac{\left(\mu_{it-1}^M\right)^j}{j!} \right\}. \tag{4}$$

Proof. The term in brackets in (4) can be shown to be finite by using the ratio test for series convergence.

We also note that since Δ_{it} has a closed form, we do not need a Newton-Raphson algorithm to evaluate it unlike in similar models for longitudinal categorical data (Heagerty, 2002; Lee and Daniels, 2007; Lee and Mercante, 2010).

2.2 *Poisson Marginalized Transition Random Effects Model*

We propose an extended model that allows both serial and long range dependence similar to that proposed in Schildcrout and Heagerty (2007) for binary data. We replace (2) in the PMTM with

$$\log(\mu_{it}^c(y_{it-1}, a_i)) = \triangle_{it} + \gamma_{it} \left\{ \log \max(\delta, y_{it-1}) - \log \mu_{it-1}^M \right\} + \sigma_i a_i, \tag{5}$$

where $\mu_{it}^c(y_{it-1}, a_i)) = E(Y_{it}|y_{it-1}, a_i)$, $\log \sigma_i = c_i^T \lambda$ with c_i being a $g \times 1$ vector of subset of x_i and λ being a $g \times 1$ coefficient of c_i, and a_i is a standard normal random variable.

Similar to calculation of $\triangle_{it}$ in the PMTM, we have the following relationship between (1) and (5), given β, γ, and λ,

$$\mu_{it}^M = \int \left\{ \sum_{j=0}^{\infty} \mu_{it}^c(j, a_i) q_{it-1j}(a_i) \right\} \phi(a_i) da_i, \tag{6}$$

where $q_{it-1j}(a_i) = P(Y_{it-1} = j|a_i)$, the calculation of which is given in the Appendix, and $\phi(a_i)$ is the standard normal probability density function. From (6), the closed form expression for Δ_{it} is given in the following theorem.

Theorem II. The parameters Δ_{it} in (5) are finite and given by

$$\triangle_{it} = x_{it}^T\beta + \gamma_{it}\log\mu_{it-1}^M - \log\left\{\int \delta^{\gamma_{it}} e^{\sigma_i a_i} q_{it-10}(a_i)\phi(a_i)da_i + \int e^{\sigma_i a_i}\sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i)\phi(a_i)da_i\right\}, \quad (7)$$

where

$$\triangle_{i1} = x_{i1}^T\beta - \frac{\sigma_i^2}{2}.$$

Proof. See the Appendix.

Again, since $\triangle_{it}$ has a closed form, we do not need a Newton-Raphson algorithm to evaluate it.

2.3 *Maximum Likelihood Estimation*

We outline a maximum likelihood algorithm for the PMTM and PMTREM next.

PMTM Let $\omega = (\beta, \alpha)$ be the vector of parameters from the model specification in (1) and (2). The log likelihood function is given by

$$\log L(\omega; y) = \sum_{i=1}^{N}\left[-\mu_{i1}^M + y_{i1}\log\left(\mu_{i1}^M\right) - \log(y_{i1}!) + \sum_{t=2}^{n_i}\{-\mu_{it}^c + y_{it}\log\mu_{it}^c - \log(y_{it}!)\}\right]. \quad (8)$$

A Fisher-scoring can be used to solve the likelihood equations,

$$\omega^{(c+1)} = \omega^{(c)} + \left[E\left(-\frac{\partial^2 \log L}{\partial\omega\partial\omega^T}\right)\right]^{-1}\frac{\partial \log L}{\partial\omega^{(c)}}.$$

The $(r+q)$-dimensional likelihood equations and information matrix are given in Appendix.

PMTREM Let $\xi = (\omega, \lambda)$. The log likelihood function is given by

$$\log L(\xi; y) = \sum_{i=1}^{N}\log\int \exp\left[y_{i1}\log\mu_{i1}^c(a_i) - \mu_{i1}^c(a_i) - \log(y_{i1}!)\right.$$
$$\left. + \sum_{t=2}^{n_i}\{y_{it}\log\mu_{it}^c(y_{it-1}, a_i) - \mu_{it}^c(y_{it-1}, a_i) - \log(y_{it}!)\}\right]\phi(a_i)da_i. \quad (9)$$

To solve the likelihood equation we use a Quasi-Newton algorithm,

$$\xi^{(c+1)} = \xi^{(c)} + \left[I(\xi^{(c)})\right]^{-1} \frac{\partial \log L}{\partial \xi^{(c)}},$$

where

$$I(\xi) = \sum_{i=1}^{N} L^{-2}(\xi; y_i) \frac{\partial L(\xi; y_i)}{\partial \xi} \frac{\partial L(\xi; y_i)}{\partial \xi^T}.$$

The $(r+q+g)$-dimensional likelihood equations and information matrix are also given in Appendix.

3. Modeling Overdispersion

In this section, we extend the PMTM and PMTREM to accommodate longitudinal count data with overdispersion (OMTM, OMTREM) not accounted for by a random intercept.

3.1 *Models*

As in Jowaheer and Sutradhar (2002), we suppose that conditional on b_i, y_{it} has Poisson distribution given by

$$f(y_{it}|b_{it}) = \frac{1}{y_{it}!} \exp\left\{y_{it}\eta_{it}(b_{it}) - \exp(\eta_{it}(b_{it}))\right\}, \tag{10}$$

$$f(y_{it}|y_{it-1}, b_{it}) = \frac{1}{y_{it}!} \exp\left\{y_{it}\eta^c_{it}(b_{it}) - \exp(\eta^c_{it}(b_{it}))\right\}, \tag{11}$$

with

$$E(Y_{it}|b_{it}) = var(Y_{it}|b_{it}) = \exp(\eta_{it}(b_{it})),$$

$$E(Y_{it}|y_{it-1}, b_{it}) = var(Y_{it}|y_{it-1}, b_{it}) = \exp(\eta^c_{it}(b_{it})),$$

and

$$\eta_{it}(b_{it}) = x_{it}^T\beta + \log(b_{it}),$$

$$\eta^c_{it}(b_{it}) = \triangle_{it} + \gamma_{it}\left\{\log\max(\delta, Y_{it-1}) - \log\theta^M_{it-1}\right\} + \log(b_{it}), \tag{12}$$

with $\theta^M_{it-1} = E(Y_{it-1})$. Next suppose that b_{it} has the gamma distribution with mean 1 and variance ν, with density

$$g(b_{it}) = \frac{1}{\Gamma(\nu^{-1})\nu^{\frac{1}{\nu}}} b_{it}^{\nu^{-1}-1} e^{-\frac{b_{it}}{\nu}}.$$

It then follows that marginally y_{it} has the negative binomial distributions given by

$$f(y_{it}) = \frac{\Gamma(\nu^{-1} + y_{it})}{\Gamma(\nu^{-1}) y_{it}!} \left(\frac{1}{1+\nu\theta^M_{it}} \right)^{\nu^{-1}} \left(\frac{\nu\theta^M_{it}}{1+\nu\theta^M_{it}} \right)^{y_{it}}, \tag{13}$$

and the distribution of y_{it} conditional on y_{it-1} also has a negative binomial distribution,

$$f(y_{it}|y_{it-1}) = \frac{\Gamma(\nu^{-1} + y_{it})}{\Gamma(\nu^{-1}) y_{it}!} \left(\frac{1}{1+\nu\theta^c_{it}} \right)^{\nu^{-1}} \left(\frac{\nu\theta^c_{it}}{1+\nu\theta^c_{it}} \right)^{y_{it}}, \tag{14}$$

which accommodates the overdispersion (indexed by ν). More specifically, under (13) and (14),

$$\begin{aligned}
&E(Y_{it}) = \theta^M_{it} = \exp(x_{it}^T\beta), \quad var(Y_{it}) = \theta^M_{it} + \nu\left(\theta^M_{it}\right)^2, \\
&E(Y_{it}|y_{it-1}) = \theta^c_{it} = \exp\left\{\triangle_{it} + \gamma_{it}\left(\log\max(\delta, Y_{it-1}) - \log\theta^M_{it-1}\right)\right\}, \qquad (15)\\
&var(Y_{it}|y_{it-1}) = \theta^c_{it} + \nu{\theta^c_{it}}^2.
\end{aligned}$$

Similar to (3), we calculate $\triangle$ using the following relationship given β, α, and ν,

$$\theta^M_{it} = \sum_{j=0} \theta^c_{it} P(Y_{it-1} = j). \tag{16}$$

From (16), we have a closed form of $\triangle_{it}$ given in the following Corollary.

Corollary I. The parameters Δ_{it} in (12) are finite and given by

$$\begin{aligned}
\triangle_{it} \;=\; & x_{it}^T\beta + \gamma_{it}\log\theta^M_{it-1} + \nu^{-1}\log\left(1+\nu\theta^M_{it-1}\right) \\
& -\log\left\{\delta^{\gamma_{it}} + \sum_{j=1}^{\infty} j^{\gamma_{it}} \frac{\Gamma(\nu^{-1}+j)}{\Gamma(\nu^{-1})j!}\left(\frac{\nu\theta^M_{it-1}}{1+\nu\theta^M_{it-1}}\right)^j\right\}.
\end{aligned}$$

Proof. Proof is similar to the proof of Theorem I.

Similar to the PMTREM, we can also extend the OMTM to accommodate long-term and serial dependence (OMTREM). Instead of (15), we use the following conditional mean,

$$E(Y_{it}|y_{it-1}, a_i) = \exp\left\{\triangle_{it} + \gamma_{it}\left(\log\max(\delta, y_{it-1}) - \log\theta^M_{it-1}\right) + \sigma_i a_i\right\} \stackrel{\text{let}}{=} \theta^c_{it}(y_{it-1}, a_i), \quad (17)$$

$$E(Y_{i1}|a_i) = \exp\left\{\triangle_{i1} + \sigma_i a_i\right\} \stackrel{\text{let}}{=} \theta^c_{i1}(a_i), \quad (18)$$

where $\log\sigma_i = c_i^T\lambda$ and $a_i \sim N(0,1)$. Then we have the following relationships

$$\theta^M_{i1} = \int \theta^c_{i1}(a_i)\phi(a_i)da_i, \quad (19)$$

$$\theta^M_{it} = \int \left\{\sum_{j=0}^{\infty} \theta^c_{it}(j, a_i)p_{it-1j}(a_i)\right\}\phi(a_i)da_i, \quad (20)$$

where $p_{it-1j}(a_i) = P(Y_{it-1} = j|a_i)$, the calculation of which is given in the Appendix.

$\triangle_{it}$ is a function of parameters (β, α, λ, ν). Given β, α, λ, and ν, a closed form expression for $\triangle_{it}$ can be found in the following Corollary.

Corollary II. The parameters Δ_{it} in (5) are finite and given by

$$\triangle_{it} = x_{it}^T\beta + \gamma_{it}\log\theta^M_{it-1} - \log\left[\int e^{\sigma_i a_i}p_{it-10}(a_i)\phi(a_i) + \int\sum_{j=1}^{\infty} j^{\gamma_{it}}e^{\sigma_i a_i}p_{it-1j}(a_i)\phi(a_i)da_i\right]. \quad (21)$$

where

$$\triangle_{i1} = x_{i1}^T\beta - \frac{\sigma_i^2}{2}.$$

Proof. Proof is similar to the proof of Theorem II.

3.2 *Maximum Likelihood Estimation*

A maximum likelihood algorithm for these models is given next.

OMTM Let $\psi = (\beta, \alpha, \nu)$. The log likelihood function is given by,

$$\log L(\psi; y) = \sum_{i=1}^{N}\left[\sum_{l=0}^{y_{i1}-1}\log(1+\nu l) + y_{i1}\log\left(\theta^M_{i1}\right) - \left(y_{i1}+\nu^{-1}\right)\log\left(1+\nu\theta^M_{i1}\right) - \log(y_{i1}!)\right.$$
$$\left. + \sum_{t=2}^{n_i}\left\{\sum_{l=0}^{y_{it}-1}\log(1+\nu l) + y_{it}\log(\theta^c_{it}) - \left(y_{it}+\nu^{-1}\right)\log(1+\nu\theta^c_{it}) - \log(y_{it}!)\right\}\right].$$

A Fisher-scoring algorithm to calculate the maximum likelihood estimator for the negative binomial dispersion parameter has been investigated in previous work (Lawless, 1987; Piegorsch, 1990). Lawless (1987) introduced a Fisher-scoring algorithm for independent negative binomial regression model. Piegorsch (1990) presented details for likelihood estimation for the dispersion parameter from a negative binomial distribution. The Fisher-scoring method can be used to solve the likelihood equations via

$$\psi^{(c+1)} = \psi^{(c)} + \left[E\left(-\frac{\partial^2 \log L}{\partial\psi\partial\psi^T}\right)\right]^{-1} \frac{\partial \log L}{\partial\psi^{(c)}}.$$

The $(r+q+1)$-dimensional likelihood equations and details on the Fisher-scoring algorithm used to solve the likelihood equations can be found in Appendix.

OMTREM Let $\zeta = (\beta, \alpha, \nu, \lambda)$. The log likelihood function is given by,

$$\begin{aligned} \log L(\zeta; y) &= \sum_{i=1}^{N} \log \int \exp\left[\sum_{l=0}^{y_{i1}-1} \log(1+\nu l) + y_{i1}\log\theta_{i1}^c(a_i) - (y_{i1}+\nu^{-1})\log\left(1+\nu\theta_{i1}^c(a_i)\right) - \log(y_{i1}!)\right. \\ &+ \left.\sum_{t=2}^{n_i}\left\{\sum_{l=0}^{y_{it}-1} \log(1+\nu l) + y_{it}\log\theta_{it}^c(y_{it-1}, a_i) - (y_{it}+\nu^{-1})\log\left(1+\nu\theta_{it}^c(y_{it-1}, a_i)\right) - \log(y_{it}!)\right\}\right] \phi(a_i)da_i, \end{aligned}$$

We use a Quasi-Newton algorithm to solve the likelihood equation,

$$\zeta^{(c+1)} = \zeta^{(c)} + \left[I(\zeta^{(c)})\right]^{-1} \frac{\partial \log L}{\partial\zeta^{(c)}},$$

where

$$I(\zeta) = \sum_{i=1}^{N} L^{-2}(\zeta; y_i) \frac{\partial L(\zeta; y_i)}{\partial\zeta} \frac{\partial L(\zeta; y_i)}{\partial\zeta^T}.$$

The $(r+q+1+g)$-dimensional likelihood equations and information matrix are also given in Appendix.

4. Simulation Study

We conducted simulations to examine the bias of marginal mean parameters in small samples and the robustness of inference on them to misspecification of the dependence in the presence of no missing data and under MAR missingness.

4.1 *PMTM*

We first simulated longitudinal count data under a PMTM. Covariates were time, group (2 levels), and interaction between time and group. The PMTM was specified as

$$\log(\mu_{it}^{M}) = \beta_{0k} + \beta_1 \cdot \text{group}_i + \beta_2 \cdot \text{time}_{it} + \beta_3 \text{group}_i \times \text{time}_{it},$$

$$\log(\mu_{it}^{c}) = \triangle_{it} + \gamma_{it} \left\{ \log \max(0.5, y_{it-1}) - \log \mu_{it-1}^{M} \right\},$$

$$\gamma_{it} = \alpha_0 + \alpha_1 \text{time}_{it}; \quad \beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (1.0, -0.1, 0.1, 0.1), \quad \alpha = (\alpha_0, \alpha_1) = (0.3, -1.5),$$

where $t = 1, \cdots, 20$, $\text{time}_{it} = (t-10)/10$, and group_i equals 0 or 1 with probability 0.5. For simulating under MAR dropout, we consider the following dropout model,

$$\text{logit} P(\text{dropout} = t | \text{dropout} \geq t) = \begin{cases} -5.0 + a y_{it-1}, & \text{if group}_i = 0 \ ; \\ -5.0 + b y_{it-1}, & \text{if group}_i = 1. \end{cases}$$

where $(a, b) = (0.6, 0.4)$ and $(0.06, 0.04)$ for high and low dropout rates, respectively. We simulated 500 data sets each with sample sizes of 20 and 100. We then fit three GEE models with independent (GEE-I), exchangeable (GEE-E), and AR(1) (GEE-A) working correlation matrices, a PMTM assuming $\gamma_{it} = \alpha_0$ (dependence mis-specified) (PMTM-W), and true model (PMTM-T).

Table 1 presents the average of point estimates and the square root of mean square error (MSE) of the marginal mean parameters when there was no missing data. In all models, the estimates were essentially unbiased for large sample (95% Monte Carlo intervals contained the true values) except for the estimates of coefficients of time in the GEEs. However, these biases were small. In the small sample specifically, the estimates were also essentially unbiased except for small bias in the estimates of the intercept in the GEE.

In the presence of MAR dropout, we considered two scenarios with high (50%) and low dropout (15%) percentages (see Table 2). In PMTM-T and PMTM-W, the estimates were essentially unbiased for both small and large sample (95% confidence intervals contained all true values) in the presence of the low dropout rate. However, the estimates in GEE had large bias for the coefficients of time. In the setting of a high dropout rate, only the estimates in PMTM-T were essentially unbiased. The root mean squared error ($RMSE = \sqrt{\text{MSE}}$) of the coefficients of Time and Group$*$Time for PMTM-T were smaller than those for GEE Models in the dropout case with considerable bias for the GEE. Thus, these simulations emphasize the importance of correctly specifying the dependence in the *presence of missing data.*

4.2 *PMTREM*

We also performed similar simulations under an PMTREM. We generated 500 simulated datasets from the PMTREM. The model had the marginal means with $(\beta_0, \beta_1, \beta_2, \beta_3) = (1.00, -0.10, 0.10, 0.10)$, conditional means with $(\alpha_0, \alpha_1) = (0.2, -0.1)$, and standard deviation with $(\exp(\lambda)) = e^{-0.5} = 0.61)$. We then fit three GEE models (GEE-I, GEE-E, GEE-A), and the true model (PMTREM).

Table 3 presents the results for the model when there was no missing data. For large sample ($N = 100$), the estimates in all the models were essentially unbiased (95% confidence intervals contained all true values). In the small sample, the estimates were also essentially unbiased except for the intercept in the GEE as with the previous simulation under the PMTM.

In the presence of MAR dropout (Table 4), the estimates were essentially unbiased in both small and large samples (95% Monte Carlo intervals contained all true values) in the presence of the low dropout rate, whereas the estimates for the GEE were biased for the intercept for small sample case. In the presence of a high dropout rate, only the estimates in PMTREM were essentially unbiased. The RMSEs for PMTREM were smaller than those for GEE models, especially in the high dropout case with considerable bias for the GEE.

5. Analysis of Two Clinical Trials

5.1 *Falls Prevention Study*

This was a group randomized trial conducted on sixteen 25-bed general medical-surgical nursing unit at Methodist-University Hospital, a 652-bed urban community hospital in Memphis, Tennessee. As mentioned in the introduction, eight nursing units were randomized to receive an intervention to increase alarm use and the other eight units utilized existing nursing care methods to minimize falls.

The 26 month study included an eight month 'observation' period before the intervention started and an 18 month 'study' period when the intervention was implemented. The primary outcome measure was the number of patient falls in each nursing unit during each month, which were ascertained by nurse-managers as well as hospital adverse event reports. A fall was defined as a sudden, unintentional change in position coming to rest on the ground or other lower level (Morse, 2009).

The primary goal was to assess whether the intervention was effective. The secondary goal was to identify unit-level factors (such as staffing) and patient level factors (medication use and case mix) associated with patient falls. As is common in hospitals, two units unexpectedly closed during the 16 month period of the study. We assume this missingness is ignorable.

The base model to determine the effectiveness of the intervention included intervention (TRT; $\text{TRT}_i = 1$ for Intervention group or 0 for control group;), $\text{PERIOD}_i = 1$ for the study period (after the first eight month) 0 for the first eight month observation period, and their interaction. Patient days for each unit during each month was included as an offset in all the models, ($\text{offset}_{it} = \text{PD}/1000$).

The covariates were time-varying (monthly) and exogenous. Unit level covariates were the monthly levels of staffing (registered nurses (RN), licensed practical nurses (LPN), and nursing assistants (NA)). The patient-level covariates included number of TeenCare (medicaid) patient days (TCARE), number of medicare patient days (MCARE), number of other insurance days (OTHINS), number of male patient days (MALE), number of non-Caucasian patient days (NWHITE), number of patient days over age 75 (PD75), and proportion of patients days from billing with psychotropic drug use (PROP-PSY)).

We fit the models proposed in Sections 2 and 3 (PMTM, PMTREM, OMTM, OMTREM) along with generalized linear mixed models with random effects. Using the likelihood ratio test (LRT) for nested models or a penalized model selection criterion, the Akaike Information Criteria (AIC) (Akaike, 1974), we determined the best fitting model. We first fit the base model to assess the impact of the intervention (Table 5). There were no significant differences among the falls rate by period or intervention. The best fitting models were OMTREM-1 (Maximized loglikelihood=-753.613; AIC=1521.226) and the negative binomial random effects model (Maximized loglikelihood=-753.951; AIC=1519.902). It appears there was minimal serial correlation after accounting for overdispersion via a negative binomial model with a random intercept.

We then fit the models including the base model terms and the unit level covariates (Table 6) and the base model terms with aggregated patient level covariates (Table 7). In all models, there were no significant coefficients of covariates.

In summary, the intervention, patient-level covariates aggregated at the unit-level, and unit-level covariates were not significantly associated with patient falls. A negative binomial model that allowed for between unit heterogeneity via a random intercept appeared to fit the data best (as assessed by AIC). The inclusion of the random intercept made the variability explained by the number of falls the previous month no longer significant; note that this also implies that there was no significant serial correlation.

5.2 *Epileptic Seizure Study*

We also use our models to analyze the epileptic seizure data, first reported in a paper by Thall and Vail (1990) and also analyzed in Breslow and Clayton (1993) and Booth et al. (2003). The study had 59 subjects with 4 visits per subject. Thall and Vail (1990) used a negative binomial model for analyzing the overdispersed data, but their approach used random effects, whereas we assume the dependence model with Markovian structure (2) and (15).

Subject 207 had what appeared to be very unusual data because both this subject's baseline and study-period numbers of seizures were huge and much larger than any other subject. We excluded this subject for our analysis like some previously published analyses (Thall and Vail, 1990; Diggle et al., 2002).

To examine treatment differences in epileptic seizure occurrence, we included treatment (Arm=0 for placebo; =1 for progabide), baseline seizure rate(log(seizure/4)), the age of the subject i (log(age)), visit ($(t-2.5)/5$ for $t = 1, 2, 3, 4$), and an interaction between treatment and baseline seizure rate. The AIC and likelihood ratio tests were again used as model selection criteria.

Each Fisher-scoring step on a Pentium with a 1.6GHz processor took about 1 and 4 seconds for the PMTM and the OMTM, respectively. In addition, using good initial values based on fitting an independent Poisson model in standard software results in a minimal number of iterations until convergence. We did not use the random effects models proposed, PMTREM and OMTREM or a negative binomial GLMM (using Proc

NLMIXED in SAS) due to convergence problems with only four visit points.

We fit and compared nine models. One was an PMTM, four were OMTM's, two were semiparametric models fit using GEE, and the remaining one was Poisson generalized linear mixed model with a random intercept (P-GLMM). The two semiparametric models, P-GEE and NB-GEE were GEE's with a Poisson and negative binomial distribution, respectively, and an AR(1) working correlation matrix. PMTM and OMTM-1 were the simplest models, with time homogeneous dependence, $\gamma_{it} = \alpha_0$. OMTM-2 allowed the dependence coefficients to depend on Treatment, $\gamma_{it} = \alpha_0 + \alpha_1 \times \text{Treatment}_i$. OMTM-3 and OMTM-4, respectively, had dependence parameters depending on baseline seizure rate, $\gamma_{it} = \alpha_0 + \alpha_2 \times \text{Baseline}_i$ and Age, $\gamma_{it} = \alpha_0 + \alpha_1 \times \text{Age}_i$, respectively. For OMTM-5, we had dependence parameters depending on Visit, $\gamma_{it} = \alpha_0 + \alpha_1 \times \text{Visit}_{it}$.

Table 8 presents maximum likelihood estimates and GEE estimates. Comparison of deviances for OMTM-1, OMTM-2, OMTM-3, OMTM-4, and OMTM-5, some of which are nested, yields $\triangle D_{12} = 2 \times (613.788 - 613.713) = 0.150$ ($p-$value$= 0.70$ on 1 d.f.), $\triangle D_{13} = 2 \times (613.788 - 613.491) = 0.594$ ($p-$value$= 0.44$ on 1 d.f.), $\triangle D_{14} = 2 \times (613.788 - 612.487) = 2.602$ ($p-$value$= 0.11$ on 1 d.f.), and $\triangle D_{15} = 2 \times (613.788 - 613.764) = 0.048$ ($p-$value$= 0.83$ on 1 d.f.). These comparison indicated that OMTM-1 was the best fitting model among OMTM's. The AIC for PMTM, P-GLMM, and OMTM-1 which were not nested were 1475.396, 1305.910, and 1243.576 indicating that OMTM-1 was the best fitting model. Although the likelihood ratio test of OMTM-1 vs. OMTM-4 did not reject the simpler model (OMTM-1), OMTM-4 did have the smallest AIC. In this model, the coefficient (α_3) of age in the dependence parameters was significant (0.72, SE=0.32, p-value=0.02); the Markov dependence varied with age. This indicates that the number of seizures at the previous visit had a larger impact on the number of seizures at the current visit for older subjects than for younger subjects; i.e., there was more persistence of seizures in older subjects. Likelihood based models, like those proposed here, easily allow dependence to depend on covariates. In addition, the large value of $\hat{\nu} = 0.29$ (p-value< 0.0001) confirms that the seizure counts data were overdispersed. The overdispersion affects the estimates of marginal mean parameters. The estimates of coefficient of treatment under the negative binomial model was not significant (p-value$=$ 0.055) unlike under the Poisson model (p-value$= 0.01$). The estimates of coefficients of Base and Age were significant (p-value< 0.0001 for Base; 0.01 for Age) and these indicated that, as baseline (age) increased, it was likely that epileptics will have more seizures.

We also note that the GEE marginal mean estimates in P-GEE (NB-GEE) were similar to those in PMTM (OMTM) because the working correlation matrix was AR(1) and this data set was complete (no dropout).

6. Conclusion

For the falls data, despite the unexpected closure of two nursing units in the intervention arm of the study, we conclude that there was no effect on the primary endpoint (falls). Because certain demographic and staffing variables may alter the risk of falling at the nursing unit level, we included several of these factors, which were available from hospital sources.

For the epileptic seizure study, we conclude that epileptics who had more baseline seizures and who were older had more seizures. We also found that the number of seizures at the previous visit had a larger impact on the number of seizures at the current visit for older subjects than for younger subjects (persistence); this was a new finding for this study.

We have proposed two marginalized transition models for longitudinal count data that directly model the marginal mean as a function of covariates while accounting for the serial correlation via a Markovian structure and longer term dependence via random effects. Unlike the first order marginalized transition models for longitudinal binary or ordinal/nominal data (Heagerty, 2002; Lee and Daniels, 2007; Lee and Mercante, 2010), orthogonality of marginal mean and dependence parameters no longer holds.

Parameter estimation was based on maximum likelihood estimation using a Fisher-scoring method for PMTM (OMTM) and a Quasi-Newton method for PMTREM (OMTREM), respectively. We also note the attractive feature that the conditional intercepts, Δ_{it} have a closed form unlike in similar models for longitudinal categorical data (Heagerty, 2002; Lee and Daniels, 2007; Lee and Mercante, 2010). We are working on making an R package to fit these models.

Simulation studies indicated that the ML estimates for the PMTM were essentially unbiased regardless of specification of dependence models when there was no missing data. However, marginal mean parameter estimates were less robust to the dependence model being incorrectly specified in PMTM under MAR missingness. In PMTREM, the marginal mean estimates were also essentially unbiased both in the complete data and in the presence of MAR dropout. This was consistent with previous studies (Heagerty, 2002; Lee and Daniels, 2007; Lee and Mercante, 2010).

Calculations and analyses in this paper were based on a 1st order PMTM. Extension to higher order is also possible. However, such models create many more dependence parameters and more complex computing and constraints. Extension to more flexible distributional specification of the random effects is also possible. This work is ongoing.

Acknowledgments

This work was supported by the following grants: NIH R01-CA 85295 (Daniels), NIH R01-AG025285 (Shorr, Daniels), NIH R01 AG033005 (Shorr, Daniels, and Lee), and NIH P30 AG028740 (Daniels).

Appendix

Proof of Theorem II

We will need the following lemma to prove Theorem II.

Lemma For any random variable X and real number $r > 0$,

1. If $E|X|^r < \infty$, then $P(|X| \geq t) = o(t^{-r})$ as $t \to \infty$.

2. If $P(|X| \geq t) = O(t^{-s})$ as $t \to \infty$, then $E|X|^r < \infty$ for all $r < s$.

Proof To prove the first part of Lemma, we use the following result: if $E|X| < \infty$, then $P(|X| \geq t) = o(t^{-1})$ as $t \to \infty$. This implies $P(|X|^r \geq t^r) = o(t^{-r})$ for $r > 0$ as $t^r \to \infty$. Then we have

$$t^r P(|X| \geq t) = t^r P(|X|^r \geq t^r) \to 0 \text{ as } t \to \infty.$$

This implies that $P(|X| \geq t) = o(t^{-t})$.

For the second part of Lemma, we have $P(|X| \geq t) = O(t^{-s})$ for $s > 0$ from the result 1 in Lemma. Then there exist a positive M such that $t^s P(|X| \geq t) < M$. Then we have

$$\begin{aligned} E|X|^r &= r \int_0^\infty t^{r-1} P(|X| \geq t) dt \\ &= r \int_0^\infty t^{r-s-1} \frac{P(|X| \geq t)}{t^{-s}} dt \\ &\leq r \int_0^\infty t^{r-s-1} M dt \\ &= rM \frac{t^{r-s}}{r-s} |_0^\infty = 0. \end{aligned}$$

Proof of Theorem II. We first prove the integrals in (7) are finite. The integrals in (7) can be reexpressed as

$$\begin{aligned} & \int \left[\sum_{j=0}^{\infty} \exp\left\{ \gamma_{it} \left(\log \max(\delta, j) - \log \mu_{it-1}^M \right) + \sigma_i a_i \right\} q_{it-1j}(a_i) \right] \phi(a_i) da_i \\ = \; & \left(\mu_{it-1}^M \right)^{-\gamma_{it}} \left\{ \int \delta^{\gamma_{it}} e^{\sigma_i a_i} q_{it-10}(a_i) \phi(a_i) da_i + \int e^{\sigma_i a_i} \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i) \phi(a_i) da_i \right\}. \quad (22) \end{aligned}$$

For the first integral in (22), since $0 \leq q_{it-10} \leq 1$, we have

$$\int e^{\sigma_i a_i} q_{it-10}(a_i)\phi(a_i)da_i \leq \int e^{\sigma_i a_i}\phi(a_i)da_i = \exp\left(\frac{\sigma_i^2}{2}\right).$$

For the second integral in (22), let $a_i^* = \arg_{a_i}\max\, q_{it-1j}(a_i)$. Then we have

$$\int e^{\sigma_i a_i} \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i)\phi(a_i)da_i \leq \int e^{\sigma_i a_i} \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i^*)\phi(a_i)da_i. \tag{23}$$

Now we claim

$$s_t(a_i^*) \stackrel{\text{let}}{=} \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i^*) < \infty.$$

For $t = 2$,

$$s_2(a_i^*) = \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{i1j}(a_i^*).$$

Since $q_{i1j}(a_i^*)$ is the probability mass function of the Poisson random variable with mean $\mu_{i1}^c(a_i^*)$, $s_2(a_i^*)$ is convergent by the ratio test for series convergence.

For $t \geq 3$ and $\gamma_{it} \leq 0$, since $j^{\gamma_{it}} \leq 1$ for $j = 1, 2, \ldots$, we have

$$s(a_i^*) \leq \sum_{j=1}^{\infty} q_{it-1j}(a_i^*) = 1.$$

Therefore, the series is convergent. For $t \geq 3$ and $\gamma_{it} > 0$, we claim that

$$P(Y_{it} \geq c|a_i^*) = O(c^{-s}) \text{ for all } s > 0.$$

To show this,

$$\begin{aligned}
& \lim_{c\to\infty} \frac{P(Y_{it} \geq c|a_i^*)}{c^{-s}} \\
= & \lim_{c\to\infty} \sum_{l=0}^{\infty} c^s P(Y_{it} \geq c|Y_{it-1} = l, a_i^*)P(Y_{it-1} = l|a_i^*) \\
= & \sum_{l=0}^{\infty} \lim_{c\to\infty} c^s P(Y_{it} \geq c|Y_{it-1} = l, a_i^*)P(Y_{it-1} = l|a_i^*). \qquad (24)
\end{aligned}$$

Now $E(Y_{it}^s|Y_{it-1} = l, a_i^*) < \infty$ since $[Y_{it}|Y_{it-1} = l, a_i^*] \sim \text{Poisson}(\mu_{it}^c(l, a_i^*))$. By the first part of the Lemma,

$$P(Y_{it} \geq c|Y_{it-1} = l, a_i^*) = o(c^{-s}) \text{ as } c \to \infty.$$

Therefore, the limit of (24) equals 0. This means that

$$P(Y_{it} \geq c|a_i^*) = o(c^{-s}).$$

Thus,

$$P(Y_{it} \geq c|a_i^*) = O(c^{-s}) \text{ for all } s > 0.$$

By the second part of the Lemma, $E(Y_{it}^{\gamma_{it}}|a_i^*) < \infty$ for all $\gamma_{it} < s$. Since s is arbitrary positive, $E(Y_{it}^{\gamma_{it}}|a_i^*) < \infty$ for $\gamma_{it} > 0$.

Detailed calculation of Fisher scoring for PMTM

$$\begin{aligned}
\frac{\partial \log L(\omega; y)}{\partial \beta} &= \sum_{i=1}^{N} \left[\left(y_{i1} - \mu_{i1}^M\right) x_{i1} + \sum_{t=2}^{n_i} \left\{ \left(y_{it} - \mu_{it}^c\right) \frac{\partial \triangle_{it}}{\partial \beta} - \left(y_{it} - \mu_{it}^c\right) \gamma_{it} x_{it-1} \right\} \right], \\
\frac{\partial \log L(\omega; y)}{\partial \alpha} &= \sum_{i=1}^{N} \sum_{t=2}^{n_i} \left[\left(y_{it} - \mu_{it}^c\right) \frac{\partial \triangle_{it}}{\partial \alpha} + \left(y_{it} - \mu_{it}^c\right) \left(\log \max(\delta, y_{it-1}) - \log \mu_{it-1}^M\right) z_{it} \right].
\end{aligned}$$

The expectations of the negative second derivatives of log likelihood are given by

$$
\begin{aligned}
E\left(-\frac{\partial^2 \log L}{\partial\beta\partial\beta^T}\right) &= \sum_{i=1}^{N}\left[\mu_{i1}^{M} x_{i1} x_{i1}^{T} + \sum_{t=2}^{n_i}\left\{E_{y_{it-1}}\left(\mu_{it}^{c}\right)\frac{\partial\triangle_{it}}{\partial\beta}\frac{\partial\triangle_{it}}{\partial\beta^T} - 2E_{y_{it-1}}(\mu_{it}^{c})\gamma_{it} x_{it-1}\frac{\partial\triangle_{it}}{\partial\beta^T}\right.\right. \\
&\quad \left.\left. + E_{y_{it-1}}\left(\mu_{it}^{c}\right)\gamma_{it}^{2} x_{it-1} x_{it-1}^{T}\right\}\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\alpha\partial\alpha^T}\right) &= \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left[E_{y_{it-1}}\left(\mu_{it}^{c}\right)\frac{\partial\triangle_{it}}{\partial\alpha}\frac{\partial\triangle_{it}}{\partial\alpha^T} + 2E_{y_{it-1}}\left\{\mu_{it}^{c}\left(\log\max(\delta, y_{it-1}) - \log\mu_{it-1}^{M}\right)\right\} z_{it}\frac{\partial\triangle_{it}}{\partial\alpha^T}\right. \\
&\quad \left. + E_{y_{it-1}}\left\{\mu_{it}^{c}\left(\log\max(\delta, y_{it-1}) - \log\mu_{it-1}^{M}\right)^{2}\right\} z_{it} z_{it}^{T}\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\alpha\partial\beta^T}\right) &= \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left[E_{y_{it-1}}\left(\mu_{it}^{c}\right)\frac{\partial\triangle_{it}}{\partial\alpha}\frac{\partial\triangle_{it}}{\partial\beta^T} + E_{y_{it-1}}\left\{\mu_{it}^{c}\left(\log\max(\delta, y_{it-1}) - \log\mu_{it-1}^{M}\right)\right\} z_{it}\frac{\partial\triangle_{it}}{\partial\beta^T}\right. \\
&\quad \left. - E_{y_{it-1}}\left(\mu_{it}^{c}\right)\gamma_{it}\frac{\partial\triangle_{it}}{\partial\alpha} x_{it-1}^{T} - E_{y_{it-1}}\left\{\mu_{it}^{c}\left(\log\max(\delta, y_{it-1}) - \log\mu_{it-1}^{M}\right)\right\}\gamma_{it} z_{it} x_{it-1}^{T}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial\triangle_{it}}{\partial\beta} &= x_{it} + \left(\gamma_{it} + \mu_{it-1}^{M}\right) x_{it-1} - \frac{\sum_{j=1}^{\infty} j^{\gamma_{it}+1}\left(\mu_{it-1}^{M}\right)^{j} x_{it-1}/j!}{\delta^{\gamma_{it}} + \sum_{j=1}^{\infty} j^{\gamma_{it}}\left(\mu_{it-1}^{M}\right)^{j}/j!}, \\
\frac{\partial\triangle_{it}}{\partial\alpha} &= \log\mu_{it-1}^{M} z_{it} - \frac{\delta^{\gamma_{it}}\log\delta z_{it} + \sum_{j=1}^{\infty} j^{\gamma_{it}}\log j\left(\mu_{it-1}^{M}\right)^{j} z_{it}/j!}{\delta^{\gamma_{it}} + \sum_{j=1}^{\infty} j^{\gamma_{it}}\left(\mu_{it-1}^{M}\right)^{j}/j!}.
\end{aligned}
$$

Note that all expectations are finite (can be checked by the ratio test for series convergence).

Detailed calculation of Quasi-Newton for PMTREM

$$
\begin{aligned}
\frac{\partial \log L(\xi; y)}{\partial \beta} &= \sum_{i=1}^{N} \frac{1}{\int L(\xi, a_i; y_i)\phi(a_i)da_i} \int L(\xi, a_i; y_i) \Bigg[(y_{i1} - \mu_{i1}^c(a_i)) \frac{\partial \triangle_{i1}}{\partial \beta} \\
&\quad + \sum_{t=2}^{n_i} \left\{ (y_{it} - \mu_{it}^c(y_{it-1}, a_i)) \left(\frac{\partial \triangle_{it}}{\partial \beta} - \gamma_{it} x_{it-1} \right) \right\} \Bigg] \phi(a_i) da_i, \\
\frac{\partial \log L(\xi; y)}{\partial \alpha} &= \sum_{i=1}^{N} \frac{1}{\int L(\xi, a_i; y_i)\phi(a_i)da_i} \int L(\xi, a_i; y_i) \Bigg[\sum_{t=2}^{n_i} (y_{it} - \mu_{it}^c(y_{it-1}, a_i)) \\
&\quad \left\{ \frac{\partial \triangle_{it}}{\partial \alpha} + \left(\log \max(\delta, y_{it-1}) - \log \mu_{it-1}^M \right) z_{it} \right\} \Bigg] \phi(a_i) da_i, \\
\frac{\partial \log L(\xi; y)}{\partial \lambda} &= \sum_{i=1}^{N} \frac{1}{\int L(\xi, a_i; y_i)\phi(a_i)da_i} \int L(\xi, a_i; y_i) \Bigg[(y_{i1} - \mu_{i1}^c(a_i)) \left(\frac{\partial \triangle_{i1}}{\partial \lambda} + \sigma_i a_i c_i \right) \\
&\quad + \sum_{t=2}^{n_i} \left\{ (y_{it} - \mu_{it}^c(y_{it-1}, a_i)) \left(\frac{\partial \triangle_{it}}{\partial \lambda} + \sigma_i a_i c_i \right) \right\} \Bigg] \phi(a_i) da_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \triangle_{i1}}{\partial \beta} &= x_{i1}, \quad \frac{\partial \triangle_{i1}}{\partial \alpha} = 0, \quad \frac{\partial \triangle_{i1}}{\partial \lambda} = -\sigma_i^2 c_i, \\
\frac{\partial \triangle_{it}}{\partial \beta} &= x_{it} + \gamma_{it} x_{it-1} - \frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \frac{\partial q_{it-10}(a_i)}{\partial \beta} + \sum_{j=0}^{\infty} j^{\gamma_{it}} \frac{\partial q_{it-1j}(a_i)}{\partial \beta} \right\} \phi(a_i) da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} q_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i) \right\} \phi(a_i) da_i}, \\
\frac{\partial \triangle_{it}}{\partial \alpha} &= \log \mu_{it-1}^M z_{it} - \frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \left(\log \delta q_{it-10}(a_i) z_{it} + \frac{\partial q_{it-10}(a_i)}{\partial \alpha} \right) \right\} \phi(a_i) da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} q_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i) \right\} \phi(a_i) da_i}, \\
\frac{\partial \triangle_{it}}{\partial \lambda} &= -\frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \left(\sigma_i a_i q_{it-10}(a_i) c_i + \frac{\partial q_{it-10}}{\partial \lambda} \right) + \left(\sigma_i a_i \sum_{j=1}^{\infty} j^{\gamma_{ij}} q_{it-1j}(a_i) c_i + \sum_{j=0}^{\infty} j^{\gamma_{it}} \frac{\partial q_{it-1j}}{\partial \lambda} \right) \right\} \phi(a_i) da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} q_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} q_{it-1j}(a_i) \right\} \phi(a_i) da_i}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
q_{i1k}(a_i) &= \frac{e^{-\mu_{i1}^c(a_i)} \left(\mu_{i1}^c(a_i) \right)^k}{k!}, \\
q_{itk}(a_i) &= \sum_{j=0}^{\infty} h_{ikj}^{(t)}(a_i) q_{it-1j}(a_i),
\end{aligned}
$$

where $h_{ikj}^{(t)}(a_i) = P(Y_{it} = k | Y_{it-1} = j, a_i) = \frac{e^{-\mu_{it}^c(j,a_i)}\left(\mu_{it}^c(j,a_i)\right)^k}{k!}$. Derivatives of $q_{itk}(a_i)$ are calculated as

$$\begin{aligned}
\frac{\partial q_{i1k}(a_i)}{\partial \beta} &= -q_{i1k}(a_i)\frac{\partial \triangle_{i1}}{\partial \beta}\left(\mu_{i1}^c(a_i) - k\right), \quad \frac{\partial q_{i1k}(a_i)}{\partial \alpha} = 0, \\
\frac{\partial q_{i1k}(a_i)}{\partial \lambda} &= -q_{i1k}(a_i)\left(\frac{\partial \triangle_{i1}}{\partial \lambda} + \sigma_i a_i c_i\right)\left(\mu_{i1}^c(a_i) - k\right), \\
\frac{\partial q_{itk}(a_i)}{\partial \beta} &= \sum_{j=0}^{\infty}\left\{\frac{\partial h_{ikj}^{(t)}(a_i)}{\partial \beta} q_{it-1j}(a_i) + h_{ikj}^{(t)}(a_i)\frac{\partial q_{it-1j}(a_i)}{\partial \alpha}\right\}, \\
\frac{\partial q_{itk}(a_i)}{\partial \beta} &= \sum_{j=0}^{\infty}\left\{\frac{\partial h_{ikj}^{(t)}(a_i)}{\partial \alpha} q_{it-1j}(a_i) + h_{ikj}^{(t)}(a_i)\frac{\partial q_{it-1j}(a_i)}{\partial \alpha}\right\}, \\
\frac{\partial q_{itk}(a_i)}{\partial \lambda} &= \sum_{j=0}^{\infty}\left\{\frac{\partial h_{ikj}^{(t)}(a_i)}{\partial \lambda} q_{it-1j}(a_i) + h_{ikj}^{(t)}(a_i)\frac{\partial q_{it-1j}(a_i)}{\partial \lambda}\right\}.
\end{aligned}$$

Detailed calculation of Fisher scoring for OMTM

$$\begin{aligned}
\frac{\partial \log L(\psi; y)}{\partial \beta} &= \sum_{i=1}^{N}\left[\left(y_{i1} - \theta_{i1}^M\right)\frac{1}{1+\nu\theta_{i1}^M}x_{i1} + \sum_{t=2}^{n_i}\left(y_{it} - \theta_{it}^c\right)\frac{1}{1+\nu\theta_{it}^c}\left(\frac{\partial \triangle_{it}}{\partial \beta} - \gamma_{it}x_{it-1}\right)\right], \\
\frac{\partial \log L(\psi; y)}{\partial \alpha} &= \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(y_{it} - \theta_{it}^c\right)\frac{1}{1+\nu\theta_{it}^c}\left\{\frac{\partial \triangle_{it}}{\partial \alpha} + \left(\log\max(\delta, y_{it-1}) - \log\theta_{it-1}^M\right)z_{it}\right\}, \\
\frac{\partial \log L(\psi; y)}{\partial \nu} &= \sum_{i=1}^{N}\left[-\sum_{l=0}^{y_{i1}-1}\frac{1}{\nu(1+\nu l)} + \left(y_{i1} - \theta_{i1}^M\right)\frac{1}{\nu(1+\nu\theta_{i1}^M)} + \frac{1}{\nu^2}\log\left(1+\nu\theta_{i1}^M\right)\right. \\
&+ \left.\sum_{t=2}^{n_i}\left\{-\sum_{l=0}^{y_{it}-1}\frac{1}{\nu(1+\nu l)} + \left(y_{it} - \theta_{it}^c\right)\left(\frac{1}{\nu(1+\nu\theta_{it}^c)} + \frac{\frac{\partial \triangle_{it}}{\partial \nu}}{1+\nu\theta_{it}^c}\right) + \frac{1}{\nu^2}\log\left(1+\nu\theta_{it}^c\right)\right\}\right].
\end{aligned}$$

The expectations of the negative second derivatives of log likelihood are given by

$$
\begin{aligned}
E\left(-\frac{\partial^2 \log L}{\partial\beta\partial\beta^T}\right) &= \sum_{i=1}^N \left[\frac{\theta_{i1}}{1+\nu\theta_{i1}} x_{i1}x_{i1}^T + \sum_{t=2}^{n_i} E_{y_{it-1}}\left(\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\left(\frac{\partial\triangle_{it}}{\partial\beta} - \gamma_{it}x_{it}\right)\left(\frac{\partial\triangle_{it}}{\partial\beta^T} - \gamma_{it}x_{it}^T\right)\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\alpha\partial\alpha^T}\right) &= \sum_{i=1}^N\sum_{t=2}^{n_i}\left[E_{y_{it-1}}\left(\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\frac{\partial\triangle_{it}}{\partial\alpha}\frac{\partial\triangle_{it}}{\partial\alpha^T} + 2E_{y_{it-1}}\left\{\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\left(\log\max(\delta, y_{it-1}) - \log\theta_{it-1}^M\right)\right\} z_{it}\frac{\partial\triangle_{it}}{\partial\alpha^T}\right. \\
&\qquad \left. + E_{y_{it-1}}\left\{\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\left(\log\max(\delta, y_{it-1}) - \log\theta_{it-1}^M\right)^2\right\} z_{it}z_{it}^T\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\alpha\partial\beta^T}\right) &= \sum_{i=1}^N\sum_{t=2}^{n_i}\left[E_{y_{it-1}}\left(\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\frac{\partial\triangle_{it}}{\partial\alpha}\left(\frac{\partial\triangle_{it}}{\partial\beta^T} - \gamma_{it}x_{it-1}^T\right)\right. \\
&\quad \left. + E_{y_{it-1}}\left\{\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\left(\log\max(\delta, y_{it-1}) - \log\theta_{it-1}^M\right)\right\} z_{it}\left(\frac{\partial\triangle_{it}}{\partial\beta^T} - \gamma_{it}x_{it-1}^T\right)\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\nu^2}\right) &= \sum_{i=1}^N\left[-E_{y_{i1}}\left(\sum_{l=0}^{y_{i1}-1}\frac{1+2\nu l}{\nu^2(1+\nu l)^2}\right) + \frac{2}{\nu^3}\log\left(1+\nu\theta_{i1}^M\right) - \frac{1}{\nu^2}\frac{\theta_{i1}^M}{1+\nu\theta_{i1}^M}\right. \\
&\quad + \sum_{t=2}^T\left\{-E_{y_{it}}\left(\sum_{l=0}^{y_{it}-1}\frac{1+2\nu l}{\nu^2(1+\nu l)^2}\right) + E_{y_{it-1}}\left(\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\left(\left(\frac{\partial\triangle_{it}}{\partial\nu}\right)^2 - \frac{1}{\nu^2}\right)\right. \\
&\quad \left.\left. + \frac{2}{\nu^3}E_{y_{it-1}}\left(\log\left(1+\nu\theta_{it}^c\right)\right)\right\}\right], \\
E\left(-\frac{\partial^2 \log L}{\partial\nu\partial\beta^T}\right) &= \sum_{i=1}^N\sum_{t=2}^{n_i} E_{y_{it-1}}\left(\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\frac{\partial\triangle_{it}}{\partial\nu}\left(\frac{\partial\triangle_{it}}{\partial\beta^T} - \gamma_{it}x_{it-1}^T\right), \\
E\left(-\frac{\partial^2 \log L}{\partial\nu\partial\alpha^T}\right) &= \sum_{i=1}^N\sum_{t=2}^{n_i}\left[E_{y_{it-1}}\left(\frac{\partial\theta_{it}^c}{1+\nu\theta_{it}^c}\right)\frac{\partial\triangle_{it}}{\partial\nu}\frac{\partial\triangle_{it}}{\partial\alpha^T}\right. \\
&\quad \left. + E_{y_{it-1}}\left\{\frac{\theta_{it}^c}{1+\nu\theta_{it}^c}\left(\log\max(\delta, y_{it-1}) - \log\theta_{it-1}^M\right)\right\}\frac{\partial\triangle_{it}}{\partial\nu}z_{it}^T\right]
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial\triangle_{it}}{\partial\beta} &= x_{it} + \gamma_{it}x_{it-1} + \frac{\theta_{it-1}x_{it-1}}{1+\nu\theta_{it-1}} - \frac{\sum_{j=1}^\infty j^{\gamma_{it}+1}\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^j\frac{1}{1+\nu\theta_{it-1}}x_{it}}{\delta^{\gamma_{it}} + \sum_{j=1}^\infty j^{\gamma_{it}}\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^{y_{it-1}}}, \\
\frac{\partial\triangle_{it}}{\partial\alpha} &= \log\theta_{it-1}^M z_{it} - \frac{\delta^{\gamma_{it}}\log\delta z_{it} + \sum_{j=1}^\infty j^{\gamma_{it}}\log j\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^j z_{it}}{\delta^{\gamma_{it}} + \sum_{j=1}^\infty j^{\gamma_{it}}\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^j}, \\
\frac{\partial\triangle_{it}}{\partial\nu} &= -\nu^{-2}\log\left(1+\nu\theta_{it-1}^M\right) + \nu^{-1}\frac{\theta_{it-1}^M}{1+\nu\theta_{it-1}^M} \\
&\quad - \frac{\sum_{j=1}^\infty j^{\gamma_{it}}\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^j\left\{-\sum_{l=0}^{j-1}\frac{1}{\nu+\nu^2 l} + j\frac{\nu^{-1}}{\left(1+\nu\theta_{it-1}^M\right)}\right\}}{\delta^{\gamma_{it}} + \sum_{j=1}^\infty j^{\gamma_{it}}\frac{\exp\left(\sum_{l=0}^{j-1}\log(\nu^{-1}+l)\right)}{j!}\left(\frac{\nu\theta_{it-1}^M}{1+\nu\theta_{it-1}^M}\right)^j}.
\end{aligned}
$$

We also note that all expectations are finite (can be checked by the ratio test for series convergence).

Detailed calculation of Quasi-Newton for OMTREM

$$
\begin{aligned}
\frac{\partial \log L(\zeta;y)}{\partial \beta} &= \sum_{i=1}^{N} \frac{1}{\int L(\zeta, a_i; y_i)\phi(a_i)da_i} \int L(\zeta, a_i; y_i) \Bigg[(y_{i1} - \theta^c_{i1}(a_i)) \frac{1}{1+\nu\theta^c_{i1}(a_i)} \frac{\partial \triangle_{i1}}{\partial \beta} \\
&\quad + \sum_{t=2}^{n_i} (y_{it} - \theta^c_{it}(y_{it-1}, a_i)) \frac{1}{1+\nu\theta^c_{it}(y_{it-1}, a_i)} \left(\frac{\partial \triangle_{it}}{\partial \beta} - \gamma_{it} x_{it} \right) \Bigg] \phi(a_i)da_i, \\
\frac{\partial \log L(\zeta;y)}{\partial \alpha} &= \sum_{i=1}^{N} \frac{1}{\int L(\zeta, a_i; y_i)\phi(a_i)da_i} \int L(\zeta, a_i; y_i) \Bigg[\sum_{t=2}^{n_i} (y_{it} - \theta^c_{it}(y_{it-1}, a_i)) \frac{1}{1+\nu\theta^c_{it}(y_{it-1}, a_i)} \\
&\quad \left\{ \frac{\partial \triangle_{it}}{\partial \beta} - \left(\log \max(\delta, y_{iy-1}) - \log \theta^M_{it-1} \right) z_{it} \right\} \Bigg] \phi(a_i)da_i, \\
\frac{\partial \log L(\zeta;y)}{\partial \nu} &= \sum_{i=1}^{N} \frac{1}{\int L(\zeta, a_i; y_i)\phi(a_i)da_i} \int L(\zeta, a_i; y_i) \Bigg[- \sum_{l=0}^{y_{i1}-1} \frac{1}{\nu(1+\nu l)} + (y_{i1} - \theta^c_{i1}(a_i)) \frac{\nu^{-1} + \frac{\partial \triangle_{i1}}{\partial \nu}}{1+\nu\theta^c_{i1}(a_i)} \\
&\quad + \frac{1}{\nu^2} \log\left(1 + \nu\theta^c_{i1}(a_i)\right) \\
&\quad + \sum_{t=2}^{n_i} \Bigg\{ - \sum_{l=0}^{y_{it}-1} \frac{1}{\nu(1+\nu l)} + (y_{it} - \theta^c_{it}(y_{it-1}, a_i)) \frac{\nu^{-1} + \frac{\partial \triangle_{it}}{\partial \nu}}{1+\nu\theta^c_{it}(y_{it-1}, a_i)} + \frac{1}{\nu^2} \log\left(1 + \nu\theta^c_{it}(y_{it-1,a_i})\right) \Bigg\} \Bigg] \phi(a_i)da_i, \\
\frac{\partial \log L(\zeta;y)}{\partial \lambda} &= \sum_{i=1}^{N} \frac{1}{\int L(\zeta, a_i; y_i)\phi(a_i)da_i} \int L(\zeta, a_i; y_i) \Bigg[(y_{i1} - \theta^c_{i1}(a_i)) \frac{1}{1+\nu\theta^c_{i1}(a_i)} \left(\frac{\partial \triangle_{i1}}{\partial \lambda} + \sigma_i a_i c_i \right) \\
&\quad + \sum_{t=2}^{n_i} (y_{it} - \theta^c_{it}(y_{it-1}, a_i)) \frac{1}{1+\nu\theta^c_{it}(y_{it-1}, a_i)} \left(\frac{\partial \triangle_{it}}{\partial \lambda} - \sigma_i a_i c_i \right) \Bigg] \phi(a_i)da_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \triangle_{i1}}{\partial \beta} &= x_{i1}, \quad \frac{\partial \triangle_{i1}}{\partial \alpha} = 0, \quad \frac{\partial \triangle_{i1}}{\partial \nu} = 0, \quad \frac{\partial \triangle_{i1}}{\partial \lambda} = -\sigma_i^2 c_i, \\
\frac{\partial \triangle_{it}}{\partial \beta} &= x_{it} + \gamma_{it} x_{it-1} - \frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \frac{\partial p_{it-10}(a_i)}{\partial \beta} + \sum_{j=1}^{\infty} j^{\gamma_{it}} \frac{\partial p_{it-1j}(a_i)}{\partial \beta} \right\} \phi(a_i)da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} p_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} p_{it-1j}(a_i) \right\} \phi(a_i)da_i}, \\
\frac{\partial \triangle_{it}}{\partial \alpha} &= \log \theta^M_{it-1} z_{it} - \frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \left(\log \delta p_{it-10}(a_i) z_{it} + \frac{\partial p_{it-10}(a_i)}{\partial \alpha} \right) + \sum_{j=1}^{\infty} j^{\gamma_{it}} \left(\log j z_{it} + \frac{\partial p_{it-1j}(a_i)}{\partial \alpha} \right) \right\} \phi(a_i)da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} p_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} p_{it-1j}(a_i) \right\} \phi(a_i)da_i}, \\
\frac{\partial \triangle_{it}}{\partial \nu} &= -\frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \frac{\partial p_{it-10}(a_i)}{\partial \nu} + \sum_{j=1}^{\infty} j^{\gamma_{it}} \frac{\partial p_{it-1j}(a_i)}{\partial \nu} \right\} \phi(a_i)da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} p_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} p_{it-1j}(a_i) \right\} \phi(a_i)da_i}, \\
\frac{\partial \triangle_{it}}{\partial \lambda} &= -\frac{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} \left(\sigma_i a_i p_{it-10}(a_i) c_i + \frac{\partial p_{it-10}(a_i)}{\partial \lambda} \right) + \sum_{j=1}^{\infty} j^{\gamma_{it}} \left(\sigma_i a_i p_{it-1j}(a_i) c_i + \frac{\partial p_{it-1j}(a_i)}{\partial \lambda} \right) \right\} \phi(a_i)da_i}{\int e^{\sigma_i a_i} \left\{ \delta^{\gamma_{it}} p_{it-10}(a_i) + \sum_{j=1}^{\infty} j^{\gamma_{it}} p_{it-1j}(a_i) \right\} \phi(a_i)da_i}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
p_{itk}(a_i) &= P(Y_{it} = k|a_i) = \sum_{j=0}^{\infty} P(Y_{it} = k|y_{it-1} = j, a_i) P(Y_{it-1} = j|a_i) \\
&= \sum_{j=0}^{\infty} f_{ikj}^{(t)}(a_i) p_{it-1j}(a_i),
\end{aligned}
$$

where

$$
\begin{aligned}
p_{i1k}(a_i) &= P(Y_{i1} = k|a_i) = \frac{\Gamma(\nu^{-1} + k)}{\Gamma(\nu^{-1})k!} \left(\frac{1}{1 + \nu\theta_{i1}^c(a_i)} \right)^{\nu^{-1}} \left(\frac{\nu\theta_{i1}^c(a_i)}{1 + \nu\theta_{i1}^c(a_i)} \right)^k, \\
f_{ikj}^{(t)}(a_i) &= P(Y_{it} = k|Y_{it-1} = j, a_i) = \frac{\Gamma(\nu^{-1} + k)}{\Gamma(\nu^{-1})k!} \left(\frac{1}{1 + \nu\theta_{it}^c(j, a_i)} \right)^{\nu^{-1}} \left(\frac{\nu\theta_{it}^c(j, a_i)}{1 + \nu\theta_{it}^c(j, a_i)} \right)^k.
\end{aligned}
$$

References

Akaike, H. (1974) A new look at the Statistical Model Identification. *IEEE Transactions Automatic Control*, 19, 716-723.

Alosh, M. (2010). Modeling longitudinal count data with dropouts. *Pharmaceutical Statistics.* 9, 35-45.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association.* 88, 125-134.

Booth, J.G., Casella, G., Friedl, H., Hobert, J.P. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, 3, 179-191.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. (2002). *Analysis of Longitudinal Data.* Oxford University Press.

Heagerty, P.J. (1999), Marginally specified logistic-normal models for longitudinal binary data, *Biometrics*, 55, 688-698.

Heagerty, P.J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data, *Biometrics*, 58, 342-351.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 88, 973–985.

Heagerty, P.J. and Zeger, S.L.(2000). Marginalized Multilevel models and likelihood inference (with discussion). *Statistical Science*, 15, 1-26.

Jowaheer, V. (2006). Model misspecification effects in clustered count data analysis. *Statistics* & *Probability Letters.* 76, 470-478.

Jowaheer, V. and Sutradhar, B. C. (2002). Analysing longitudinal count data with overdispersion. *Biometrika.* 89, 389-399.

Kaciroti, N.A., Raghunathan, T.E., Schork, M.A. and Clark, N.M. (2008). A Bayesian model for longitudinal count data with non-ignorable dropout. *Applied Statistics.* 57, 521-534.

Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics.* 15, 209-225.

Lee, K. and Daniels, M. (2007). A class of Markov models for longitudinal ordinal data. *Biometrics*, 63, 1060-1067.

Lee, K. and Daniels, M. (2008). Marginalized models for longitudinal ordinal data with application to quality of life studies. *Statistics in Medicine.* **27**; 4359–4380.

Lee, K., Joo, Y., Yoo, J. K., and Lee, J. (2009). Marginalized random effects models for multivariate longitudinal binary data. *Statistics in Medicine.* **28**, 1284–1300.

Lee, K., Kang, S., Liu, X., and Seo, D. (2010). Likelihood-based approach for analysis of longitudinal nominal data using marginalized random effects models. To appear in *Journal of Applied Statistics.*

Lee, K. and Mercante, D. (2010). Longitudinal nominal data analysis using marginalized models. *Computational Statistics* & *Data Analysis.* 54, 208-218.

Lee, K., Joo, Y., Song, J. J. and Harper D. (2011). Analysis of longitudinal zero-inflated count data using marginalized models. *Computational Statistics* & *Data Analysis.* 55, 824-837.

Li, J., Yang, X., Wu, Y., and Shoptaw, S. (2007). A random-effects Markov transition model for Poisson-distributed repeated measures with nonignorable missing values. *Statistics in Medicine*, 26, 2519-2532.

Molenberghs, G., Verbeke, G., and Demetrio, C.G.B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Life Data Analysis*, 13, 513-531.

Morse J. M. (2009). *Preventing Patient Falls,* Second Edition, New York, NY: Springer Publishing Company.

Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameters. *Biometrics*, 46, 863-867.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association.* 95, 1143-1156.

Schildcrout, J. and Heagerty, P. J. (2007). Marginalized models for moderate to long series of longitudinal binary response Data. *Biometrics.* **63**, 322331.

Shorr, R. I., Chandler, A. M., Mion, L. C., Daniels, M. J., Kessler, L. A., Miller, S. T. (2011). Effectiveness of a bed alarm system to prevent falls in hospitalized patients. Submitted.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46, 657-671.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika.* 75, 621-629.

Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics.* 44, 1019-1031.

Table 1

Bias of PMTM maximum likelihood estimators with sample sizes of 20 and 100 in the complete data . Displayed are the average regression coefficient estimates, the square root of MSE, $\sqrt{\sum_{i=1}^{500}(\theta-\hat{\theta}_i)^2/500}$, and 95% Monte Carlo confidence interval ($\bar{\beta} \pm 1.96\sqrt{var(\bar{\beta})/500}$).

	N=20										N=100									
	GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W		GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W	
Parameter	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$
Intercept	0.99	0.07	0.99	0.07	0.99	0.07	0.99	0.07	0.99	0.07	0.99	0.03	0.99	0.03	0.99	0.03	1.00	0.03	1.00	0.03
(1.00)	(0.98,0.99)		(0.98,0.99)		(0.98,0.99)		(0.98,1.00)		(0.99,1.00)		(0.99,0.99)		(0.99,0.99)		(0.99,0.99)		(1.00,1.00)		(1.00,1.00)	
Group	-0.10	0.11	-0.10	0.11	-0.10	0.11	-0.10	0.11	-0.10	0.11	-0.10	0.05	-0.10	0.05	-0.10	0.05	-0.10	0.05	-0.10	0.05
(-0.10)	(-0.11,-0.09)		(-0.11,-0.09)		(-0.11,-0.09)		(-0.11,-0.09)		(-0.11,-0.09)		(-0.10,-0.10)		(-0.10,-0.10)		(-0.10,-0.10)		(-0.11,-0.10)		(-0.10,-0.10)	
Time	0.08	0.14	0.08	0.14	0.08	0.14	0.09	0.13	0.10	0.14	0.08	0.06	0.08	0.06	0.08	0.06	0.10	0.05	0.10	0.06
(0.10)	(0.07,0.10)		(0.07,0.10)		(0.07,0.09)		(0.08,0.10)		(0.09,0.11)		(0.08,0.09)		(0.08,0.09)		(0.08,0.09)		(0.09,0.10)		(0.09,0.10)	
Group*Time	0.10	0.18	0.10	0.18	0.10	0.18	0.09	0.18	0.10	0.18	0.10	0.09	0.10	0.09	0.10	0.09	0.10	0.08	0.10	0.09
(0.10)	(0.08,0.11)		(0.08,0.11)		(0.08,0.12)		(0.08,0.11)		(0.08,0.12)		(0.09,0.11)		(0.10,0.11)		(0.10,0.11)		(0.09,0.11)		(0.10,0.11)	

Table 2

Bias of PMTM maximum likelihood estimators with sample size of 20 and 100 in the presence of MAR. Displayed are the average regression coefficient estimates, the square root of MSE, $\sqrt{\sum_{i=1}^{500}(\theta-\hat{\theta}_i)^2/500}$, and 95% Monte Carlo confidence interval ($\bar{\beta} \pm 1.96\sqrt{var(\bar{\beta})/500}$).

	High dropout										Low dropout									
$N = 20$	GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W		GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W	
Parameter	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$
Intercept	0.96	0.11	0.98	0.11	0.98	0.11	1.00	0.11	0.99	0.11	0.99	0.04	0.99	0.05	0.99	0.05	1.00	0.04	1.00	0.05
(1.00)	(0.96,0.97)		(0.97,0.99)		(0.97,0.99)		(0.99,1.01)		(0.98,0.99)		(0.99,1.00)		(0.99,1.00)		(0.99,1.00)		(1.00,1.00)		(1.00,1.00)	
Group	-0.09	0.14	-0.10	0.14	-0.10	0.14	-0.11	0.14	-0.10	0.14	-0.10	0.06	-0.10	0.06	-0.10	0.07	-0.10	0.06	-0.10	0.07
(-0.10)	(-0.10,-0.07)		(-0.11,-0.09)		(-0.11,-0.09)		(-0.12,-0.10)		(-0.11,-0.09)		(-0.11,-0.10)		(-0.11,-0.10)		(-0.11,-0.10)		(-0.11,-0.10)		(-0.11,-0.10)	
Time	0.02	0.20	0.04	0.19	0.04	0.19	0.08	0.18	0.05	0.19	0.08	0.09	0.08	0.09	0.08	0.09	0.10	0.07	0.10	0.08
(0.10)	(0.01,0.04)		(0.03,0.06)		(0.02,0.05)		(0.07,0.10)		(0.03,0.06)		(0.07,0.09)		(0.07,0.09)		(0.07,0.09)		(0.09,0.10)		(0.09,0.10)	
Group*Time	0.13	0.25	0.12	0.24	0.12	0.24	0.10	0.23	0.12	0.24	0.10	0.12	0.10	0.12	0.10	0.12	0.10	0.11	0.10	0.12
(0.10)	(0.11,0.15)		(0.10,0.14)		(0.10,0.14)		(0.08,0.12)		(0.10,0.14)		(0.09,0.11)		(0.09,0.11)		(0.09,0.11)		(0.09,0.11)		(0.09,0.11)	
$N = 100$	GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W		GEE-I		GEE-E		GEE-A		PMTM-T		PMTM-W	
Parameter	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$	Mean	$\sqrt{MSE}$
Intercept	0.96	0.06	0.98	0.05	0.98	0.05	0.99	0.04	0.98	0.05	0.99	0.03	0.99	0.03	0.99	0.03	1.00	0.03	1.00	0.03
(1.00)	(0.95,0.96)		(0.97,0.98)		(0.97,0.98)		(0.99,1.00)		(0.97,0.98)		(0.99,0.99)		(0.99,0.99)		(0.99,0.99)		(1.00,1.00)		(1.00,1.00)	
Group	-0.08	0.06	-0.09	0.06	-0.09	0.06	-0.10	0.06	-0.09	0.06	-0.10	0.05	-0.10	0.05	-0.10	0.05	-0.10	0.05	-0.10	0.05
(-0.10)	(-0.08,-0.07)		(-0.09,-0.08)		(-0.09,-0.08)		(-0.10,-0.09)		(-0.09,-0.08)		(-0.10,-0.10)		(-0.10,-0.10)		(-0.10,-0.10)		(-0.10,-0.10)		(-0.10,-0.10)	
Time	0.02	0.11	0.04	0.09	0.03	0.10	0.09	0.08	0.04	0.09	0.08	0.06	0.08	0.06	0.08	0.06	0.10	0.05	0.10	0.06
(0.10)	(0.01,0.03)		(0.04,0.05)		(0.03,0.04)		(0.09,0.10)		(0.04,0.05)		(0.08,0.09)		(0.08,0.09)		(0.08,0.09)		(0.10,0.10)		(0.09,0.11)	
Group*Time	0.14	0.11	0.13	0.11	0.13	0.11	0.11	0.10	0.14	0.11	0.10	0.09	0.10	0.09	0.10	0.09	0.10	0.08	0.10	0.09
(0.10)	(0.13,0.15)		(0.12,0.14)		(0.12,0.14)		(0.10,0.12)		(0.13,0.15)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)		(0.09,0.11)		(0.10,0.11)	

Table 3

Bias of PMTREM maximum likelihood estimators with sample sizes of 20 and 100 in the complete data . Displayed are the average regression coefficient estimates, the square root of MSE, $\sqrt{\sum_{i=1}^{500}(\theta-\hat{\theta}_i)^2/500}$, *and 95% Monte Carlo confidence interval* $(\bar{\beta} \pm 1.96\sqrt{var(\bar{\beta})/500})$.

	N=20								N=100							
	GEE-I		GEE-E		GEE-A		PMTREM		GEE-I		GEE-E		GEE-A		PMTREM	
Parameter	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$
Intercept	0.98	0.22	0.98	0.22	0.97	0.23	0.98	0.20	1.00	0.11	1.00	0.11	1.00	0.11	1.00	0.09
(1.00)	(0.96,1.00)		(0.96,1.00)		(0.95,0.99)		(0.97,1.00)		(0.99,1.01)		(0.99,1.01)		(1.00,1.01)		(1.00,1.01)	
Group	-0.11	0.30	-0.11	0.30	-0.10	0.31	-0.11	0.29	-0.11	0.15	-0.11	0.15	-0.11	0.15	-0.10	0.14
(-0.10)	(-0.14,-0.08)		(-0.14,-0.08)		(-0.13,-0.07)		(-0.14,-0.08)		(-0.12,-0.10)		(-0.12,-0.10)		(-0.12,-0.10)		(-0.11,-0.09)	
Time	0.11	0.08	0.11	0.08	0.10	0.13	0.10	0.08	0.11	0.03	0.11	0.03	0.11	0.06	0.10	0.03
(0.10)	(0.10,0.11)		(0.10,0.11)		(0.09,0.11)		(0.09,0.11)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)	
Group*Time	0.11	0.12	0.11	0.12	0.11	0.19	0.10	0.11	0.10	0.05	0.10	0.05	0.09	0.08	0.10	0.05
(0.10)	(0.10,0.12)		(0.10,0.12)		(0.09,0.13)		(0.09,0.11)		(0.10,0.11)		(0.10,0.11)		(0.09,0.10)		(0.10,0.11)	

Table 4

Bias of PMTREM maximum likelihood estimators with sample size of 20 and 100 in the presence of MAR. Displayed are the average regression coefficient estimates, the square root of MSE, $\sqrt{\sum_{i=1}^{500}(\theta-\hat{\theta}_i)^2/500}$, *and 95% Monte Carlo confidence interval* $(\bar{\beta} \pm 1.96\sqrt{var(\bar{\beta})/500})$.

	High dropout								Low dropout							
$N = 20$	GEE-I		GEE-E		GEE-A		PMTREM		GEE-I		GEE-E		GEE-A		PMTREM	
Parameter	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$
Intercept	0.71	0.34	0.94	0.23	0.88	0.27	0.97	0.19	0.97	0.22	0.98	0.21	0.97	0.22	0.98	0.15
(1.00)	(0.70,0.73)		(0.92,0.96)		(0.86,0.90)		(0.95,0.98)		(0.95,0.99)		(0.96,1.00)		(0.95,0.99)		(0.97,1.00)	
Group	0.04	0.30	-0.09	0.30	-0.06	0.30	-0.12	0.26	-0.08	0.32	-0.09	0.31	-0.08	0.32	-0.11	0.22
(-0.10)	(0.01,0.06)		(-0.12,-0.06)		(-0.09,-0.04)		(-0.15,-0.10)		(-0.11,-0.06)		(-0.12,-0.06)		(-0.11,-0.06)		(-0.12,-0.09)	
Time	-0.14	0.31	0.07	0.12	0.04	0.25	0.09	0.13	0.10	0.10	0.11	0.08	0.10	0.14	0.10	0.09
(0.10)	(-0.16,-0.12)		(0.06,0.08)		(-0.05,-0.02)		(0.08,0.10)		(0.09,0.11)		(0.10,0.11)		(0.09,0.12)		(0.10,0.11)	
Group*Time	0.22	0.27	0.12	0.15	0.17	0.26	0.11	0.17	0.10	0.15	0.10	0.12	0.10	0.20	0.09	0.13
(0.10)	(0.20,0.25)		(0.11,0.13)		(0.15,0.19)		(0.10,0.13)		(0.09,0.11)		(0.09,0.11)		(0.08,0.12)		(0.08,0.10)	
$N = 100$	GEE-I		GEE-E		GEE-A		PMTREM		GEE-I		GEE-E		GEE-A		PMTREM	
Parameter	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$	Mean	$\sqrt{\text{MSE}}$
Intercept	0.72	0.30	0.94	0.15	0.90	0.18	0.99	0.10	0.99	0.10	1.00	0.10	1.00	0.11	1.00	0.09
(1.00)	(0.71,0.73)		(0.93,0.95)		(0.88,0.91)		(0.98,0.10)		(0.98,1.00)		(0.99,1.01)		(0.99,1.01)		(0.99,1.01)	
Group	0.05	0.19	-0.07	0.13	-0.05	0.14	-0.10	0.14	-0.09	0.15	-0.09	0.14	-0.09	0.15	-0.10	0.13
(-0.10)	(0.04,0.06)		(-0.08,-0.06)		(-0.06,-0.04)		(-0.12,-0.09)		(-0.11,-0.08)		(-0.11,-0.09)		(-0.10,-0.08)		(-0.12,-0.09)	
Time	-0.12	0.24	0.07	0.06	-0.03	0.16	0.10	0.06	0.10	0.05	0.10	0.04	0.10	0.06	0.10	0.04
(0.10)	(-0.13,-0.11)		(0.07,0.07)		(-0.04,-0.02)		(0.09,0.10)		(0.09,0.10)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)	
Group*Time	0.21	0.15	0.11	0.07	0.16	0.13	0.10	0.07	0.10	0.06	0.10	0.05	0.10	0.09	0.10	0.05
(0.10)	(0.20,0.22)		(0.11,0.12)		(0.15,0.17)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)		(0.10,0.11)	

Table 5

*Base Models with an intercept, TRT, PERIOD, and TRT*PERIOD*

	PMTM			PMTREM			OMTM			OMTREM			GLMM	
	1	2	3	1	2	3	1	2	3	1	2	3	Poisson	N.B.
Intercept(β_0)	1.72	1.72	1.72	1.66	1.66	1.66	0.71	0.17	1.71	1.68	1.67	1.68	1.63	1.64
	(0.10)	(0.11)	(0.12)	(0.11)	(0.13)	(0.16)	(0.10)	(0.10)	(0.11)	(0.13)	(0.16)	(0.19)	(0.12)	(0.12)
TRT(β_1)	0.10	0.10	0.10	0.14	0.15	0.14	0.08	0.08	0.07	0.12	0.14	0.12	0.15	0.13
	(0.12)	(0.15)	(0.12)	(0.17)	(0.19)	(0.32)	(0.14)	(0.14)	(0.15)	(0.19)	(0.22)	(0.35)	(0.17)	(0.17)
PERIOD(β_2)	-0.14	-0.15	-0.14	-0.11	-0.10	-0.11	-0.15	-0.15	-0.15	-0.12	-0.12	-0.12	-0.11	-0.11
	(0.10)	(0.10)	(0.14)	(0.11)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.16)	(0.17)	(0.18)	(0.09)	(0.10)
TRT*PERIOD(β_3)	0.00	0.01	0.00	0.07	0.06	0.06	0.03	0.03	0.03	0.08	0.08	0.08	0.06	0.08
	(0.15)	(0.15)	(0.15)	(0.13)	(0.15)	(0.13)	(0.18)	(0.18)	(0.18)	(0.20)	(0.21)	(0.20)	(0.13)	(0.15)
Intercept(α_0)	0.17*	0.19*	0.20*	0.02	-0.01	0.04	0.16*	0.18*	0.22*	0.03	0.00	0.05		
	(0.04)	(0.06)	(0.06)	(0.05)	(0.11)	(0.11)	(0.05)	(0.07)	(0.09)	(0.06)	(0.14)	(0.13)		
TRT(α_1)		-0.05			0.06			-0.03			0.05			
		(0.08)			(0.12)			(0.10)			(0.16)			
PERIOD(α_2)			-0.05			-0.02			-0.08			-0.02		
			(0.08)			(0.19)			(0.11)			(0.24)		
Intercept(λ_0)				-1.37*	-1.36*	-1.38*				-1.42*	-1.40*	-1.43*	0.26(σ)	0.25(σ)
				(0.25)	(0.25)	(0.28)				(0.27)	(0.26)	(0.30)	(0.06)	(0.06)
Dispersion(ν)							0.13*	0.13*	0.13*	0.09*	0.08*	0.09*		0.08*
							(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.03)
Max.loglik.	-771.762	-771.699	-771.508	-758.649	-758.442	-758.616	-761.270	-761.285	-761.168	-753.613	-753.438	-753.559	-758.760	-753.951
AIC	1553.524	1555.398	1555.016	1529.298	1530.884	1531.232	1534.540	1536.570	1536.336	1521.226	1522.876	1523.118	1527.500	1519.902

* indicates significance with 95% confidence level.

Table 6

*Models with an intercept, TRT, PERIOD, TRT*PERIOD, and staffing (RN, LPN, NA)*

	PMTM		PMTREM		OMTM		OMTREM		GLMM	
	1	2	1	2	1	2	1	2	Poisson	N.B.
Intercept(β_0)	1.70	1.68	1.62	1.62	1.69	1.68	1.68	1.67	1.64	1.64
	(0.10)	(0.10)	(0.18)	(0.26)	(0.10)	(0.10)	(0.21)	(0.23)	(0.12)	(0.12)
TRT(β_1)	0.08	0.08	0.14	0.14	0.06	0.06	0.12	0.11	0.15	0.13
	(0.12)	(0.13)	(0.20)	(0.28)	(0.14)	(0.15)	(0.22)	(0.26)	(0.17)	(0.18)
PERIOD(β_2)	−0.13	−0.07	−0.12	−0.10	−0.15	−0.11	−0.13	−0.11	−0.15	−0.15
	(0.10)	(0.10)	(0.15)	(0.17)	(0.12)	(0.12)	(0.17)	(0.16)	(0.10)	(0.11)
TRT*PERIOD(β_3)	0.01	−0.03	0.07	0.03	0.04	0.00	0.01	0.07	0.10	0.10
	(0.14)	(0.15)	(0.15)	(0.16)	(0.17)	(0.17)	(0.21)	(0.20)	(0.13)	(0.15)
RN(β_4)	−0.02	−0.03	0.02	0.10	−0.02	−0.01	0.01	−0.00	0.07	0.06
	(0.06)	(0.07)	(0.10)	(0.10)	(0.07)	(0.07)	(0.10)	(0.11)	(0.08)	(0.09)
LPN(β_5)	0.07	0.08	0.04	0.06	0.06	0.08	0.04	0.05	0.06	0.06
	(0.06)	(0.06)	(0.10)	(0.11)	(0.06)	(0.06)	(0.10)	(0.11)	(0.06)	(0.07)
NA(β_6)	0.04	0.02	−0.01	−0.00	0.06	0.03	0.01	0.01	−0.03	−0.02
	(0.07)	(0.07)	(0.09)	(0.11)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)
Intercept(α_0)	0.13*	0.15*	0.02	0.04	0.15*	0.16*	0.03	0.05		
	(0.05)	(0.04)	(0.09)	(0.09)	(0.05)	(0.05)	(0.10)	(0.13)		
RN(α_1)		−0.20*		−0.08		−0.19*		−0.10		
		(0.04)		(0.14)		(0.05)		(0.12)		
Intercept(λ_0)			−1.40*	−1.50*			−1.46*	−1.59*	0.26* (σ)	0.25* (σ)
			(0.35)	(0.40)			(0.38)	(0.44)	(0.06)	(0.07)
Dispersion(ν)					0.12*	0.11*	0.08*	0.08*		0.08*
					(0.04)	(0.04)	(0.04)	(0.05)		(0.03)
Max.loglik.	−768.009	−763.214	−757.870	−756.543	−758.649	−755.093	−753.029	−750.554	−757.475	−752.968
AIC	1552.018	1544.428	1533.740	1533.086	1535.298	1530.186	1526.058	1523.108	1530.950	1523.936

* indicates significance with 95% confidence level.

Table 7

*Models with an intercept, TRT, PERIOD, TRT*PERIOD, TCARE, MCARE, OTHINS, MALE, NWHITE, PD75, and PROP-PSY*

	PMTM		PMTREM		OMTM		OMTREM		GLMM	
	1	2	1	2	1	2	1	2	Poisson	N.B.
Intercept(β_0)	1.73	1.69	1.63	1.60	1.72	1.69	1.69	1.67	1.64	1.65
	(0.10)	(0.10)	(0.21)	(0.37)	(0.10)	(0.10)	(0.16)	(0.16)	(0.12)	(0.12)
TRT(β_1)	0.11	0.14	0.15	0.18	0.08	0.10	0.12	0.15	0.15	0.13
	(0.12)	(0.12)	(0.26)	(0.26)	(0.14)	(0.14)	(0.23)	(0.23)	(0.17)	(0.17)
PERIOD(β_2)	−0.17	−0.09	−0.14	−0.09	−0.19	−0.14	−0.15	−0.11	−0.13	−0.14
	(0.11)	(0.10)	(0.28)	(0.48)	(0.13)	(0.12)	(0.23)	(0.21)	(0.10)	(0.11)
TRT*PERIOD(β_3)	−0.01	−0.07	0.06	0.01	0.02	−0.02	0.08	0.04	0.06	0.07
	(0.14)	(0.14)	(0.32)	(0.40)	(0.17)	(0.17)	(0.27)	(0.27)	(0.13)	(0.15)
TCARE(β_4)	−0.06	−0.06	−0.06	−0.05	−0.07	−0.07	−0.06	−0.06	−0.06	−0.06
	(0.06)	(0.06)	(0.12)	(0.14)	(0.07)	(0.07)	(0.13)	(0.13)	(0.06)	(0.07)
MCARE(β_5)	−0.02	−0.01	−0.08	−0.08	−0.02	−0.02	−0.07	−0.06	−0.08	−0.07
	(0.10)	(0.10)	(0.26)	(0.24)	(0.12)	(0.12)	(0.28)	(0.29)	(0.11)	(0.13)
OTHINS(β_6)	0.16*	0.17*	0.08	0.14	0.17*	0.18*	0.09	0.09	0.08	0.09
	(0.06)	(0.07)	(0.18)	(0.32)	(0.08)	(0.07)	(0.18)	(0.20)	(0.08)	(0.08)
MALE(β_7)	−0.10	−0.12	−0.03	−0.05	−0.10	−0.11	−0.04	−0.05	−0.03	−0.04
	(0.07)	(0.07)	(0.10)	(0.29)	(0.09)	(0.09)	(0.11)	(0.12)	(0.07)	(0.08)
NWHITE(β_8)	0.04	0.03	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.07
	(0.11)	(0.11)	(0.33)	(0.33)	(0.13)	(0.13)	(0.33)	(0.32)	(0.13)	(0.14)
PD75(β_9)	0.06	0.05	0.03	0.03	0.06	0.06	0.03	0.03	0.03	0.02
	(0.05)	(0.05)	(0.12)	(0.16)	(0.06)	(0.06)	(0.10)	(0.10)	(0.06)	(0.06)
PROP-PSY(β_{10})	0.02	0.03	−0.02	−0.01	0.03	0.04	−0.00	0.00	−0.01	−0.00
	(0.04)	(0.04)	(0.08)	(0.15)	(0.05)	(0.05)	(0.07)	(0.07)	(0.04)	(0.05)
Intercept(α_0)	0.13*	0.15*	0.02	0.04	0.14*	0.16*	0.03	0.04		
	(0.04)	(0.04)	(0.08)	(0.08)	(0.05)	(0.05)	(0.05)	(0.05)		
OTHINS(α_1)		−0.09*		−0.06		−0.08		−0.06		
		(0.04)		(0.29)		(0.05)		(0.06)		
Intercept(λ_0)			−1.39*	−1.46*			−1.48*	−1.51*	0.25*(σ)	0.24*(σ)
			(0.52)	(0.85)			(0.27)	(0.28)	(0.06)	(0.07)
Dispersion(ν)					0.12*	0.11*	0.08*	0.08*		0.08*
					(0.04)	(0.04)	(0.03)	(0.03)		(0.03)
Max.loglik.	−766.379	−764.742	−756.777	−755.849	−757.132	−756.017	−752.064	−751.102	−756.960	−752.437
AIC	1556.758	1555.484	1539.554	1539.698	1540.264	1540.034	1532.128	1532.204	1537.92	1530.874

* indicates significance with 95% confidence level.

Table 8

Maximum likelihood estimates for PMTM and OMTM and GEE estimates.

	Model								
	PMTM	OMTM-1	OMTM-2	OMTM-3	OMTM-4	OMTM-5	P-GEE	NB-GEE	P-GLMM
Int	−2.65 (0.56)	−1.74 (0.99)	−1.74 (0.99)	−1.52 (0.99)	−2.14 (0.98)	−1.74 (0.99)	−2.53 (0.84)	−1.67 (0.91)	−1.23 (1.16)
Trt	-0.62^* (0.25)	−0.69 (0.37)	−0.69 (0.37)	-0.69 (0.36)	−0.71 (0.37)	−0.69 (0.37)	−0.62 (0.40)	−0.69 (0.37)	−0.70 (0.42)
Base	0.94^* (0.06)	0.90^* (0.11)	0.89^* (0.11)	0.89^* (0.11)	0.86^* (0.11)	0.90^* (0.11)	0.94^* (0.09)	0.89^* (0.11)	0.88^* (0.13)
Age	0.87^* (0.16)	0.62^* (0.29)	0.62^* (0.29)	0.55 (0.29)	0.76^* (0.29)	0.62^* (0.29)	0.86^* (0.25)	0.63^* (0.27)	0.47 (0.34)
Visit	−0.07 (0.16)	-0.20^* (0.21)	−0.20 (0.21)	−0.21 (0.21)	−0.18 (0.20)	−0.20 (0.21)	−0.05 (0.04)	−0.04 (0.03)	−0.04 (0.02)
Trt*Base	0.17 (0.11)	0.20 (0.19)	0.20 (0.19)	0.20 (0.19)	0.22 (0.18)	0.20 (0.19)	0.17 (0.19)	0.20 (0.19)	0.19 (0.22)
Int	0.38^* (0.02)	0.35^* (0.07)	0.37^* (0.10)	0.14 (0.20)	−2.03 (1.06)	0.35^* (0.08)			
Trt			−0.04 (0.14)						
Base				0.11 (0.10)					
Age					0.72^* (0.32)				
Visit						0.04 (0.44)			
overdis.		0.30^* (0.05)	0.29^* (0.05)	0.30^* (0.05)	0.29^* (0.05)	0.29^* (0.05)			$0.49^*(\sigma)(0.06)$
Max. lik.	−730.698	−613.788	−613.713	−613.491	−612.487	−613.764			−645.955
AIC	1475.396	1243.576	1245.426	1244.982	1242.974	1245.528			1305.910

* indicates significance with 95% confidence level.