Two-Part Pattern-Mixture Model for Longitudinal Incomplete Semi-Continuous Toenail Data

Rajibul Islam Mian[a] and M. Tariqul Hasan[b,*]

[a]*Department of Mathematics and Statistics, University of Windsor, ON, Canada*

[b]*Department of Mathematics and Statistics, University of New Brunswick, NB, Canada*

Abstract: Longitudinal data with true zero values, known as longitudinal semi-continuous data, frequently occur in medical, environmental and biological studies. To model longitudinal semi-continuous data, two-part modelling approaches have been widely used in literature. In the first part of the two-part model, binary logistic regression is commonly used after converting the semi-continuous responses to binary responses. In the second part, the semi-continuous data are converted to positive continuous data after removing the true zero values from the responses. Although positive continuous or non-zero values tend to show a positively skewed distribution pattern, in the literature the normal distribution is commonly used to model them. Also, in longitudinal studies, data often suffer individual dropouts as they are collected overtime. In this paper, we propose a two-part pattern-mixture model to analyze longitudinal semi-continuous data with dropouts. In the proposed approach, we use pattern-mixture binary mixed models for the first part and positively continuous pattern-mixture gamma mixed models for the second part. Our approach can accommodate both subject- and time-specific correlation as well as dropout pattern. We also incorporate a computationally efficient estimation method for our models using a penalize quasi-likelihood approach. The proposed method is illustrated with an application to the longitudinal incomplete toenail data.

Keywords: Two-part model, semi-continuous data, gamma mixed model, dropout, Toenail data.

1. INTRODUCTION

1.1. Motivating Example

In the past few decades, medical science has received greater attention than many other disciplines. In medical research, one of the main concentrations is to develop and compare the effectiveness as well as the safety of the drugs for treating various diseases. This work is motivated by the toenail onychomycosis study, where the main interest is to compare the efficiency of two antifungal compounds. Toenail onychomycosis is a very common (6-8% among adult persons) toenail disease, which causes almost half of all toenail related abnormalities. Usually antifungal compounds were used as treatment of this kind of toenail disease. Terbinafine (also known as Lamisil) and Itraconazole (two antifungal compounds) have reduced the duration of treatments to 3 months as compared to other treatments [1]. The aim of the study was to compare the efficacy and safety features of these two antifungal compounds. Patients with the toenail disease of age 18 or older were included in the study. Three hundred and seventy eight patients were randomized and received a box of study medications. Each daily dose contained either a 250 mg Terbinafine tablet (250 mg/day dose) and placebo capsules or two 100 mg Itraconazole capsules (200 mg/day dose) and a placebo tablet. Patients under study were asked to take two capsules and a tablet daily after dinner for 3 months. One target nail was selected for each study patient at the beginning of the study and the sample was taken from that nail. For every study patient, the follow-up records were obtained in 0, 1, 2, 3, 6, 9 and 12 months. Measurements of unaffected nail lengths were considered responses, which were taken at the baseline and every follow-up visit. The response zero indicates that the length of the unaffected part of the nail is zero (i.e. fully affected nail) and the non-zero response indicates the length of unaffected part of nail. Toenail data is longitudinal semi-continuous data as the responses include the exact zeros and positive non-zero values. The toenail data was first analyzed by Backer *et al*. [1] by using the Mantel-Haenzel test, Breslow-Day test, two-sample binomial test and some exploratory analysis. Komáreck and Lesaffre [2] used a general linear mixed model to analyze toenail data considering the normal distribution assumption for the responses with one random effect. Toenail data was also analyzed by using various simple statistical tools [3-5].

1.2. Background of the Study

To analyze longitudinal semi-continuous data, two-part mixed model approaches have been proposed by various authors [6-8]. Most of these approaches use a combination of a binary mixed model and a Gaussian mixed model to analyze the zero and non-zero parts of the responses respectively. Our explanatory analysis of

*Address corresponding to this author at the Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada; Tel: 1-506-458-7367; Fax: 1-506-453-4705; E-mail: thasan@unb.ca

E-ISSN: 1929-6029/12

the toenail data shows the positive skewed nature for the non-zero part (Figures **1** and **2**) and hence using a Gaussian mixed model may not be appropriate to analyze this data set [9]. Toenail data set also suffer by the individual dropouts which commonly occur in the longitudinal studies, as the responses are collected repeatedly over time. So far, to the best of our knowledge, accommodating a dropout mechanism for longitudinal semi-continuous data has received little attention in the literature.

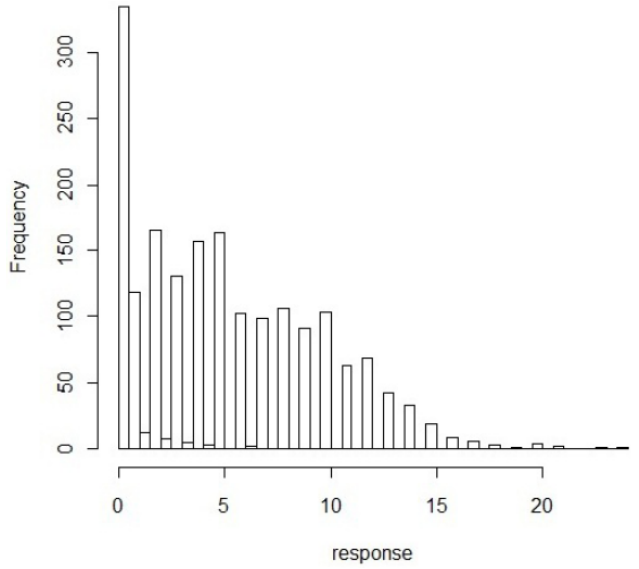

Figure 1: Response frequency plot for complete toenail data.

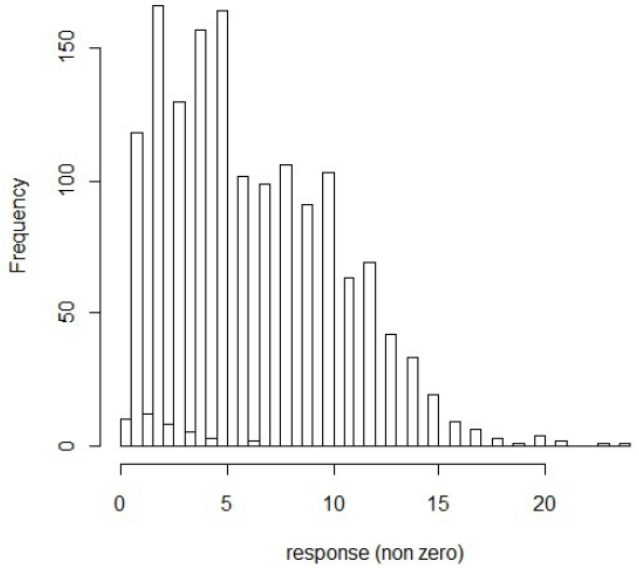

Figure 2: Response frequency plot for non-zero part of the toenail data.

1.3. Proposed Method

In this paper, we introduce a two-part model to analyze semi-continuous incomplete toenail data. In the first part of the two-part model, we use the pattern-mixture mixed effect logistic regression model after converting the semi-continuous responses to binary responses. In the second part, unlike other approaches, we will use pattern-mixture gamma mixed models to model the non-zero parts of the data. Our proposed approach will be able to accommodate subject-specific and time-specific variations. Our pattern-mixture approach will also be able to incorporate the dropout pattern of the toenail data. We introduce the proposed two-part pattern-mixture mixed model in Section 2 and their moment structures in Section 3. We discuss a penalized quasi-likelihood approach for parameter estimation in Section 4. The analysis of the toenail data is presented in Section 5.

2. TWO-PART PATTERN-MIXTURE MIXED MODEL

In this section, we discuss a two-part pattern-mixture mixed model for incomplete longitudinal semi-continuous data. Let Y_{ij} represents the semi-continuous response i.e. the length of unaffected part of toenail for the ith $(i=1,2,...,m)$ subject at jth $(j=1,2,...,n_i)$ occasion. In the next subsection, we first discuss the pattern-mixture binary logistic regression model.

2.1. Pattern-Mixture Mixed Effect Logistic Regression Model

In the first part, we rearrange semi-continuous responses Y_{ij} to the binary responses Y_{ij}^* where Y_{ij}^* = 1 if $Y_{ij}=0$ or 0 otherwise. Then the binary response vector can be expressed as $\boldsymbol{Y}^*=(Y_{11}^*,...,Y_{1n_1}^*,...,Y_{m1}^*,...,Y_{mn_m}^*)'$. Let U_i^* be the subject-specific random effect for the response of the i th subject and V_{ij}^* be the time-specific random effect at the j th time point of the i th subject. These random effects U_i^* and V_{ij}^* will be able to incorporate subject-specific and time-specific variations in the binary responses, respectively. Our mixed effect logistic regression model is developed under the following four assumptions:

Assumption 1: Subject-specific random effects $U_1^*,...,U_i^*,...,U_m^*$ are independently and identically distributed as normal with mean zero and variance $\sigma_{u^*}^2$, i.e. $U_i^* \sim N(0,\sigma_{u^*}^2)$

Assumption 2: Time-specific random effects $V_{11}^*,...,V_{1n_1}^*,...,V_{m1}^*,...,V_{mn_m}^*$ are independently and identically distributed as normal with mean zero and variance $\sigma_{v^*}^2$, i.e. $V_{ij}^* \sim N(0,\sigma_{v^*}^2)$.

Assumption 3: The random effects U_i^* and V_{ij}^* are independent.

Assumption 4: Conditional on the random effects $\boldsymbol{U}^* = (U_1^*,...,U_i^*,...,U_m^*)'$ and $\boldsymbol{V}^* = (V_{11}^*,...,V_{1n_1}^*,...,V_{m1}^*,...,V_{mn_m}^*)'$, the response Y_{ij}^* independently follows a Bernoulli distribution with parameter π_{ij}, i.e.

$$Y_{ij}^* | \boldsymbol{U}^*, \boldsymbol{V}^* \sim Bernoulli(\pi_{ij}) \ \ where \ \ \pi_{ij} = Pr(Y_{ij}^* = 1). \qquad (1)$$

In (1), the logit of π_{ij} can be expressed using the following link function as

$$logit(\pi_{ij}) = \beta_0 + \beta_1 T_{ij}^* + \beta_2 (T_{ij}^*)^2 + \beta_3 G_{ij}^* + \sum_{l=1}^{6} \beta_{3+l} Z_{ijl} + U_i^* + V_{ij}^*, \quad (2)$$

where $\beta = (\beta_0, \beta_1, ..., \beta_9)'$ represents the vector of the regression parameters. In (2), T_{ij}^*, $(T_{ij}^*)^2$ represents the linear and the quadratic evolution of time respectively and G_{ij}^* is the treatment group indicator variable for the i th subject at j th occasion. Note that in toenail data, $G_{ij}^* = 1$ indicates treatment group A, i.e. patients taking Terbinafine and $G_{ij}^* = 0$ indicates treatment group B, i.e. patients taking Itraconazole. In (2), Z_{ijl} (for $l = 1,2,...,6$) indicates the dummy variables for the dropout pattern, which can be expressed as in Table **1** [10]. The model discussed in this section will be used in Section 4 to analyze the binary part of the two-part model.

2.2. Pattern-Mixture Mixed Effect Gamma Regression Model

In this section we will discuss the second part of the model using a positively skewed pattern-mixture gamma mixed model. To do that we first discuss the gamma regression model in Section 2.2.1. In Section 2.2.2, we will discuss the pattern-mixture gamma mixed models.

2.2.1. Gamma Regression Model

Recall that to model the non-zero responses of the two-part model, the Gaussian distribution is commonly used in literature. In Figure **2**, we showed that the non-zero responses are positively skewed. As argued by Anderson *et. al.* [9] the gamma distribution is a more logical choice for skewed and non-negative responses. Therefore, in the second part, we use gamma regression to model the response variable. Note that the gamma distribution is a positively skewed continuous distribution which is formed by using two parameters: a shape parameter θ_1 and scale parameter θ_2 (known as Gamma (θ_1, θ_2)). The density function of Gamma (θ_1, θ_2) can be expressed as

$$f(y) = \frac{1}{\Gamma\theta_1 \theta_2^{\theta_1}} y^{\theta_1 - 1} \exp(-y/\theta_2); \ y > 0.$$

with $E(Y) = \theta_1\theta_2$ and $var(Y) = \theta_1\theta_2^2$. In Figure **3**, we present some examples of gamma distributions plotted for various values of θ_1 and θ_2. To do that, we consider $\theta_1 = 2.5$ and $\theta_2 = 0.3, 0.5$ and 0.8. Note that in this paper, we use a gamma regression model where the mean of the responses is a link function of the covariates and corresponding unknown parameters. In the next section, we will discuss the gamma regression model for the semi-continuous longitudinal incomplete data.

2.2.2. Pattern-Mixture Mixed Effect Gamma Regression Model

In the second part of the model, we reorganize the semi-continuous responses Y_{ij} to the positive continuous responses Y_{ij}^{**} by ignoring the exact zero responses. Let Y_{ij}^{**} represents the positive continuous

Table 1: Dummy Codes by Pattern of Dropout

Pattern Groups	Weeks of follow ups							Dummy Variables					
	0	4	8	12	24	36	48	Z_{ij1}	Z_{ij2}	Z_{ij3}	Z_{ij4}	Z_{ij5}	Z_{ij6}
1	O	D	D	D	D	D	D	1	0	0	0	0	0
2	O	O	D	D	D	D	D	0	1	0	0	0	0
3	O	O	O	D	D	D	D	0	0	1	0	0	0
4	O	O	O	O	D	D	D	0	0	0	1	0	0
5	O	O	O	O	O	D	D	0	0	0	0	1	0
6	O	O	O	O	O	O	D	0	0	0	0	0	1
7	O	O	O	O	O	O	O	0	0	0	0	0	0

D: Dropout; O: Observed.

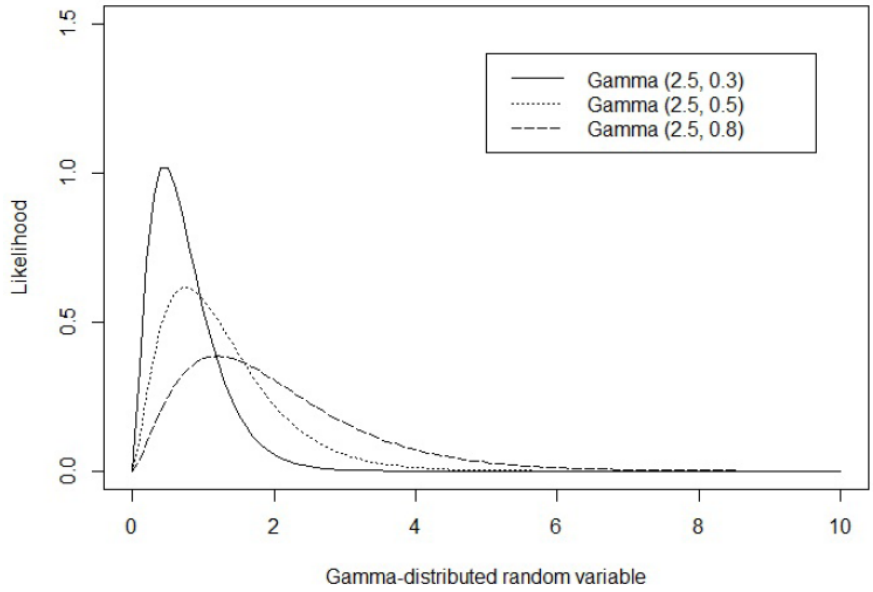

Figure 3: Likelihood plot for Gamma distribution for various θ_1 and θ_2.

response at the j th $(j=1,2,...,h_i)$ time point of the i th $(i=1,2,...,m)$ independent subject. Note that, h_i, the repeated number of responses for the i th subject under the gamma mixed model, is less than or equal to the total number repeated responses n_i. So, the response vector can be expressed as $\boldsymbol{Y}^{**}=(Y_{11}^{**},...,Y_{1h_1}^{**},...,Y_{m1}^{**},...,Y_{mh_m}^{**})'$. Similar to mixed effect logistic regression, U_i^{**} is the subject-specific random effect for the response of the i th subject and V_{ij}^{**} is the time-specific random effect at the j th time point of the i th subject. The gamma regression model is developed under the following four assumptions:

Assumption 1. Subject-specific random effects $U_1^{**},...,U_i^{**},...,U_m^{**}$ are independently and identically distributed as normal with mean zero and variance $\sigma_{u^{**}}^2$, i.e. $U_i^{**}\sim N(0,\sigma_{u^{**}}^2)$.

Assumption 2. Time-specific random effects $V_{11}^{**},...,V_{1h_1}^{**},...,V_{m1}^{**},...,V_{mh_m}^{**}$ are independently and identically distributed as normal with mean zero and variance $\sigma_{v^{**}}^2$, i.e. $V_{ij}^{**}\sim N(0,\sigma_{v^{**}}^2)$.

Assumption 3. The random effects U_i^{**} and V_{ij}^{**} are independent.

Assumption 4. Conditional on the random effects $\boldsymbol{U}^{**}=(U_1^{**},...,U_i^{**},...,U_m^{**})'$ and $\boldsymbol{V}^{**}=(V_{11}^{**},...,V_{1h_1}^{**},...,V_{m1}^{**},...,V_{mh_m}^{**})'$ the response Y_{ij}^{**} independently follows a Gamma distribution with parameters α and λ_{ij} i.e.,

$$Y_{ij}^{**}\,|\boldsymbol{U}^{**},\boldsymbol{V}^{**}\sim Gamma(\alpha,\lambda_{ij}), \tag{3}$$

where α and λ_{ij}, represents the shape and the scale parameters respectively. The conditional mean of Y_{ij}^{**} given the random effects, can be expressed as $\mu_{ij}=\alpha\lambda_{ij}$, where μ_{ij} can be defined using various link functions. The link functions we use in this paper are as follows

$$Inverse\ link,\ g(\mu_{ij})=\frac{1}{\lambda_{ij}}=\delta_0+\delta_1T_{ij}^{**}+\delta_2(T_{ij}^{**})^2+\delta_3G_{ij}^{**}$$
$$+\sum_{l=1}^{6}\delta_{3+l}Z_{ijl}+U_i^{**}+V_{ij}^{**}$$

$$Log\ link,\ g(\mu_{ij})=log(\lambda_{ij})=\delta_0+\delta_1T_{ij}^{**}+\delta_2(T_{ij}^{**})^2+\delta_3G_{ij}^{**}$$
$$+\sum_{l=1}^{6}\delta_{3+l}Z_{ijl}+U_i^{**}+V_{ij}^{**}$$

$$Identity\ link,\ g(\mu_{ij})=\lambda_{ij}=\delta_0+\delta_1T_{ij}^{**}+\delta_2(T_{ij}^{**})^2+\delta_3G_{ij}^{**}$$
$$+\sum_{l=1}^{6}\delta_{3+l}Z_{ijl}+U_i^{**}+V_{ij}^{**}$$

By using the inverse link function, λ_{ij} can written as

$$\lambda_{ij}^{-1}=\delta_0+\delta_1T_{ij}^{**}+\delta_2(T_{ij}^{**})^2+\delta_3G_{ij}^{**}$$
$$+\sum_{l=1}^{6}\delta_{3+l}Z_{ijl}+U_i^{**}+V_{ij}^{**}, \tag{4}$$

where $\delta=(\delta_0,\delta_1,...,\delta_9)'$ represents the vector of the regression parameters under pattern-mixture gamma mixed model. Similar to logistic mixed models, T_{ij}^{**},

$(T_{ij}^{**})^2$ and G_{ij}^{**} represents the covariates for the i th subject at j th occasion. In (4), Z_{ijl} (for $l = 1,2,\ldots,6$) indicates the dummy variables for the dropout pattern as discussed in Section 2.1. In the next section we will discuss estimation of the model parameters.

3. MOMENT STRUCTURE

In this section we will discuss the moment structure of the mixed effect logistic and gamma regression models discussed in the previous section. First we briefly present the mixed effect logistic regression models. Let $\boldsymbol{X}_{ij}^{*} = (1, T_{ij}^{*}, (T_{ij}^{*})^2, G_{ij}^{*}, Z_{ij1}, \ldots, Z_{ij6})'$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_9)'$. The unconditional expectation of the response Y_{ij}^{*} can be expressed as

$$E(Y_{ij}^{*}) = EE(Y_{ij}^{*} | \boldsymbol{U}^{*}, \boldsymbol{V}^{*}) = E(\pi_{ij})$$
$$where \quad \pi_{ij} = \frac{\exp(X_{ij}^{*'}\beta + U_i^{*} + V_{ij}^{*})}{1 + \exp(X_{ij}^{*'}\beta + U_i^{*} + V_{ij}^{*})}. \tag{5}$$

The unconditional variance of the response Y_{ij}^{*} has the form

$$\begin{aligned} var(Y_{ij}^{*}) &= E[var(Y_{ij}^{*} | \boldsymbol{U}^{*}, \boldsymbol{V}^{*})] + var[E(Y_{ij}^{*} | \boldsymbol{U}^{*}, \boldsymbol{V}^{*})] \\ &= E[\pi_{ij}(1 - \pi_{ij})] + var(\pi_{ij}). \end{aligned} \tag{6}$$

Similarly the unconditional covariance of the responses Y_{ij}^{*} and $Y_{i'j'}^{*}$ can be expressed as

$$cov(Y_{ij}^{*}, Y_{i'j'}^{*}) = \begin{cases} cov(Y_{ij}^{*}, Y_{ij'}^{*}) & if\ i = i'\ and\ j \neq j' \\ 0 & otherwise \end{cases}. \tag{7}$$

After some algebra, it is possible to show that $\text{cov}(Y_{ij}^{*}, Y_{ij'}^{*}) = \text{cov}(\pi_{ij}, \pi_{ij'})$. It is cumbersome to evaluate the exact mathematical expression of the correlation structures. To get an approximate idea about the correlation structure between response variables $\text{corr}(Y_{ij}^{*}, Y_{ij'}^{*})$, we will find the correlation $\text{corr}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*})$, as $\text{corr}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*})$ is a function of $\text{corr}(Y_{ij}^{*}, Y_{ij'}^{*})$. This will help us understand the correlation structure within each subject. The correlation, $\text{corr}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*})$ can be written as

$$\text{corr}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*}) = \frac{\text{cov}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*})}{\sqrt{\text{var}(U_i^{*} + V_{ij}^{*}) \times \text{var}(U_i^{*} + V_{ij'}^{*})}},$$

where $\text{cov}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*}) = \sigma_{u^{*}}^2$ and $\text{var}(U_i^{*} + V_{ij}^{*}) = \sigma_{u^{*}}^2 + \sigma_{v^{*}}^2$, which implies

$$\text{corr}(U_i^{*} + V_{ij}^{*}, U_i^{*} + V_{ij'}^{*}) = \frac{\sigma_{u^{*}}^2}{\sigma_{u^{*}}^2 + \sigma_{v^{*}}^2}. \tag{8}$$

So, we could assume that for the same subject, correlation between different time points approximately follows an exchangeable correlation structure. A similar exchangeable correlation structure also appears for the responses under the mixed effect gamma regression models which are not presented in this paper.

4. ESTIMATION OF PARAMETERS

To estimate the regression and random effect parameters, we use the penalized quasi-likelihood technique [11] by maximizing the joint densities. For the mixed effect logistic regression model, the conditional density of the response given the random effects is

$$f(Y_{ij}^{*} | \boldsymbol{U}^{*}, \boldsymbol{V}^{*}, \beta) \sim Bernoulli(\pi_{ij}). \tag{9}$$

The marginal likelihood function can be evaluated by integrating the joint likelihood, which can be written as

$$L_1(\beta, \sigma_{U^{*}}^2, \sigma_{V^{*}}^2) = \int\int f(Y_{ij}^{*}, \boldsymbol{U}^{*}, \boldsymbol{V}^{*}; \beta, \sigma_{U^{*}}^2, \sigma_{V^{*}}^2) dU^{*} dV^{*}. \tag{10}$$

To estimate the regression and random effect parameters, one needs to solve the following estimating equations simultaneously:

$$\begin{aligned} &\frac{\partial L_1(\beta, \sigma_{U^{*}}^2, \sigma_{V^{*}}^2)}{\partial \beta} = 0; \quad \frac{\partial L_1(\beta, \sigma_{U^{*}}^2, \sigma_{V^{*}}^2)}{\partial \sigma_{U^{*}}^2} = 0 \\ &\text{and} \quad \frac{\partial L_1(\beta, \sigma_{U^{*}}^2, \sigma_{V^{*}}^2)}{\partial \sigma_{V^{*}}^2} = 0. \end{aligned} \tag{11}$$

Similarly, for the mixed effect gamma regression model, the conditional density of the response given the random effects is

$$f(Y_{ij}^{**} | \boldsymbol{U}^{**}, \boldsymbol{V}^{**}, \delta) \sim Gamma(\alpha, \lambda_{ij}), \tag{12}$$

and the marginal likelihood function can be written as

$$L_2(\delta, \sigma_{U^{**}}^2, \sigma_{V^{**}}^2) = \int\int f(Y_{ij}^{**}, \boldsymbol{U}^{**}, \boldsymbol{V}^{**}; \delta, \sigma_{U^{**}}^2, \sigma_{V^{**}}^2) dU^{**} dV^{**}. \tag{13}$$

Similar to the binary logistic regression model, we estimate the regression and random effect parameters for the gamma regression model by solving the following estimating equations simultaneously:

$$\begin{aligned} &\frac{\partial L_2(\delta, \sigma_{U^{**}}^2, \sigma_{V^{**}}^2)}{\partial \delta} = 0; \quad \frac{\partial L_2(\delta, \sigma_{U^{**}}^2, \sigma_{V^{**}}^2)}{\partial \sigma_{U^{**}}^2} = 0 \\ &\text{and} \quad \frac{\partial L_2(\delta, \sigma_{U^{**}}^2, \sigma_{V^{**}}^2)}{\partial \sigma_{V^{**}}^2} = 0. \end{aligned} \tag{14}$$

Table 2: Observed Number of Dropouts at Various Visit of the Study

Months	Treat A(1)	Treat B(0)	Total	Dropout
0	150	148	298	-
1	149	142	291	7
2	146	138	284	7
3	140	131	271	13
6	131	124	255	16
9	120	109	229	26
12	118	108	226	3

It is highly cumbersome to evaluate (11) and (14). We have used the $glmmPQL$ function of the $MASS$ package in the statistical software R to estimate the regression and random effect parameters of the model.

5. DATA ANALYSIS

In this section, we analyze the semi-continuous toenail data by using the proposed two-part pattern-mixture models. As mentioned earlier toenail data were first analyzed by Backer *et al*. [1], where the response variable indicates the length of unaffected part of the nail. Unaffected nail length is measured from the nail bed to the infected part of the nail. The data only include patients for whom the measurements were taken from any one of the two big toenails. This constraint reduces the data to 150 subjects in group A (patients taking Terbinafine) and 148 subjects in group B (patients taking Itraconazole). We took the measurement of unaffected nail lengths as the response variable, and time, quadratic evolution of time (timesq) and treatment as covariates. At first, we show the overall picture of the dropout situation of the toenail data in Table **2** which presents the visit wise dropout counts. We analyze the toenail data using the pattern-mixture mixed effect logistic and gamma regression model for the first and second part respectively. Our initial analysis using a binary logistic mixed model shows that the dropout patterns are insignificant. Our mixed effect gamma regression models using log, identity and inverse link functions, also show similar conclusions about the dropout patterns. Then we reanalyzed the toenail data by simultaneously removing the pattern groups from the data set as they have found to be insignificant.

In the first part, we apply the mixed effect logistic regression model presented in section 2.1 without considering the pattern groups. This will reduces the logit in (2) to $logit(\pi_{ij}) = \beta_0 + \beta_1 T_{ij}^* + \beta_2 (T_{ij}^*)^2 + \beta_3 G_{ij}^* + U_i^* + V_{ij}^*$. Our data analysis results, presented in Table **3**, show that both covariates, time and quadratic evolution of time (timesq) are significant. From the results, we find that the covariate time has negative significant effect, which means fully affected nail lengths (zero response) would decrease as time increases. Although our results also show that the covariate timesq has positive significant effect, this indicates the fully affected nail lengths would increase at a lower rate, in comparison with less affected nail lengths. The result of our data analysis also suggests that the two treatments have no significant difference in effectiveness. In the mixed effect logistic regression part of the model, we found that after considering the covariates, the remaining grouping or clustering effects at the time level is larger than that of the subject level as the estimates of $\sigma_{u^*}^2$ and $\sigma_{v^*}^2$ are 2.4849 and 4.8686, respectively. These subject- and time-specific variations indicate that there is still a large amount of variation left beyond that which can be captured by the model. As the time-specific

Table 3: Mixed Effect Logistic Regression Model Results

Variable Name	Estimate	Std. Error	t value	P-Value
Intercept	-0.3554	0.1713	-2.0748	0.0381
Time	-1.0829	0.0487	-22.2022	0.0000
Timesq	0.0769	0.0040	19.0453	0.0000
Treatment	-0.1733	0.2118	-0.8183	0.4138
σ_u^2*	2.4849			
σ_v^2*	4.8686			

variation is larger than the subject-specific variation, one can argue that the time-specific random effect is important for consideration in the model and the improvement of the completely affected nails varies by the duration of treatment used.

In the second part of the model, we use the mixed effect gamma regression model discussed in Section 2.2. To do that we have used identity, inverse and log links. As the results are similar for all three links, we only report the result for the log link in Table **4**. For the second part of the model, we have considered the non-zero responses of all the persons with available number of occasions as data. By using the mixed effect gamma regression model we found both covariates, time and the quadratic evolution of time (timesq) are significant at the 5% level. The results we have found in our analyses show that the less affected nail lengths will increase as time increases and the less affected nail lengths will decrease as quadratic evolution of time increases. Note that as in the binary part of the analysis, the fully affected nail lengths would increase at a lower rate, in comparison with less affected nail lengths. Moreover, as in the first part of the model, the second part of the model shows that the two treatments have no significant difference in effectiveness as the indicator variable for the treatments was found insignificant. We also found that after considering the covariates, the remaining grouping or clustering effects at the subject level and the time level estimates of σ^2_{u**} and σ^2_{v**} are 0.2457 and 0.1849 respectively. The small values of σ^2_{u**} and σ^2_{v**} indicate that we can capture most of the subject- and time-specific variations by the model.

6. CONCLUSION

We have proposed a two-part pattern-mixture mixed model to analyze longitudinal semi-continuous toenail data. Two-part models facilitate the separate analysis for the zero and non-zero parts of the data set. In each part of the two-part models we incorporated the subject-specific as well as the time-specific random effects which help us describe the variability that arose from the subjects and the occasions respectively. The pattern-mixture approach to each part of the two-part models also helps us recognize the effects and impacts of dropout in the data set.

As mentioned earlier, the toenail data was first analyzed by Backer *et al*. [1] by using the Mantel-Haenzel test, Breslow-Day test, two-sample binomial test and some exploratory analysis. The main purpose of their study was to compare the two antifungal compounds, Terbinafine and Itraconazole, and they found Terbinafine showed better performance than Itraconazole [1]. But our results indicate that there is no significant difference between the effectiveness of the two compounds. Verbeke and Molenberghs [12] incorporated the dropout patterns in the mixed effect model to analyze the toenail data. In their models they considered only subject-specific random effects and unstructured correlation between outcomes within subjects. They also assumed the normal distribution assumption for the responses. These studies [1, 12] did not point out one of the important characteristics of the toenail data: this data set is positively skewed with excessive number of zeros. Considering both treatment groups and all occasions, 17.5296% zeros are available in the data set as response. Consideration of these extra zeros may affect the distributional assumption of the response variable of the data set.

Though we obtained similar results as Verbeke and Molenberghs [12], we have applied a different modeling aspect than others. We have dealt with extra zero features of the data by using the two part models. In our analysis, we have considered exchangeable correlation structure between outcomes within subjects and indicator variable for the treatment groups. That means the pairwise correlation between the responses is the same for various time points. Moreover, our analysis captured the subject-specific as well as the time-specific random effects for the mixed effect model. We have also managed the dropout patterns of the data by using the pattern-mixture approach to each part of the two part models. These extra features help our model to be more appropriate for the toenail data.

Table 4: Mixed Effect Gamma Regression Model Results for Log Link

Variable Name	Estimate	Std. Error	t value	P-Value
Intercept	0.7715	0.0519	14.8401	0.0000
Time	0.2679	0.0122	21.8702	0.0000
Timesq	-0.0138	0.0009	-14.1302	0.0000
Treatment	0.0493	0.0644	0.7648	0.4450
σ^2_u**	0.2457			
σ^2_v**	0.1849			

ACKOWLEDGEMENTS

This research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. Authors would like to thank the editor, associate editor and referee for the helpful comments, which have served to greatly improve the quality of the paper. Authors also would like to thank Drs. Renjun Ma and Gary Sneddon for their help.

REFERENCES

[1] Backer MDe, Vroery CDe, Lesaffre E, Scheys I, Keyser PDe. Twelve weeks of continuous oral therapy for toenail onychomycosis caused by dermatophytes: a double-blind comparative trial of Terbinafine 250 mg/day versus Itraconazole 200 mg/day. J Am Acad Dermatol 1998; 38: 57-63. http://dx.doi.org/10.1016/S0190-9622(98)70486-4

[2] Komáreck A, Lesaffre E. Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. Comput Statist Data Anal 2008; 52: 3441-58. http://dx.doi.org/10.1016/j.csda.2007.10.024

[3] Gianni C. Update on antifungal therapy with Terbinafine. Giornale Italiano di Dermatologia e Venereologia 2010; 145: 415-24. http://ukpmc.ac.uk/abstract/MED/20461049

[4] Newland JG, Abdel-Rahman SM. Update on Terbinafine with a focus on dermatophytoses. Clin Cosmetic Investig Dermatol 2009; 2: 49-63. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047923/

[5] Zanini M. Surgical treatment of onychomycosis. Medicina Cutanea Ibero-Latino-Americana 2009; 37: 67-8. http://www.medcutan-ila.org/articulos/2009/1/pdf/mc371l.pdf

[6] Lu SE, Lin Y, Shih WC. Analyzing Excessive No Changes in Clinical Trials with Clustered Data. Biometrics 2004; 60: 257-67. http://www.ncbi.nlm.nih.gov/pubmed/15032797 http://dx.doi.org/10.1111/j.0006-341X.2004.00155.x

[7] Olsen MK, Schafer JL. A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data. J Am Stat Assoc 2011; 96: 730-45. http://www.tandfonline.com/doi/abs/10.1198/016214501753168389 http://dx.doi.org/10.1198/016214501753168389

[8] Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. U.S. National Library of Medicine 2002. doi: 10.1191/0962280202sm291ra

[9] Anderson CJ, Verkuilen J, Johnson T. Applied generalized linear mixed models: continuous and discrete data. New York: Springer 2010.

[10] Hasan MT, Sneddon G, Ma R. Pattern-mixture zero-inflated mixed models for longitudinal unbalanced count data with excessive zeros. Biometr J 2009; 51: 946-60. http://dx.doi.org/10.1002/bimj.200900093

[11] Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J of the Amer Stat Assoc 1993; 88: 9-25. http://dx.doi.org/10.2307/2290687

[12] Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer-Verlag 2000.

Received on 04-10-2012 Accepted on 26-11-2012 Published on 21-12-2012

http://dx.doi.org/10.6000/1929-6029.2012.01.02.05