Analysis of Mixed Discrete and Heavy Tailed Longitudinal Data with Non-random Missingness Using Stochastic Variants of the EM Algorithm

Abdallah S. A. Yaseen[*]

The National Centre for Social and Criminological Research, Cairo, Egypt
*Corresponding author: aswanwater@gmail.com

Abstract Interstitial cystitis (IC) is a chronic inflammatory condition that results in recurring discomfort or pain in the bladder and the surrounding pelvic region. In interstitial cystitis data base (ICDB) cohort study, the main target is to determine the influence of covariates, such as the demographic clinical characteristics of patients, on the longitudinal outcomes including the pain score (p), urinary urgency (u) and urinary frequency (f) which are three main indices reflecting IC symptoms. The ICDB data are mixed (discrete and continuous) longitudinal data. In longitudinal studies the continuous response may be non-normal, heavy tailed for example. The analysis of mixed longitudinal data is challenging due to several inherent features: (1) more than one outcome are followed for each subject over a period of time. (2) The longitudinal outcomes are subject to missingness that may be missing not at random (MNAR). This article proposes the analysis of mixed discrete and heavy tailed longitudinal outcomes subject to MNAR missingness using two different alternative algorithms. The continuous outcome is assumed to follow non-normal heavy tailed distribution. The proposed methodology is an extension of [1] and [2]. The proposed techniques are applied to Interstitial Cystitis data. Also, three simulation studies are conducted to validate the proposed techniques.

Keywords: stochastic expectation maximization, parametric fractional imputation, interstitial cystitis, longitudinal data, maximum likelihood, missing data

Cite This Article: Abdallah S. A. Yaseen, "Analysis of Mixed Discrete and Heavy Tailed Longitudinal Data with Non-random Missingness Using Stochastic Variants of the EM Algorithm." *American Journal of Applied Mathematics and Statistics*, vol. 10, no. 1 (2022): 28-38. doi: 10.12691/ajams-10-1-5.

1. Introduction

Longitudinal data are comprised of repeated observations of one or more outcomes and a set of covariates for many subjects. A special type of longitudinal data is the bivariate or multivariate longitudinal data where more than one outcome is followed and jointly studied. Multivariate longitudinal data are useful to investigate the joint evolution of these responses over time. The scale (discrete/continuous) of multiple longitudinal outcomes is crucial in multivariate longitudinal analysis. Several approaches are used to analyze mixed discrete and continuous longitudinal data. A general approach for jointly modeling mixed discrete and continuous outcomes is to use the location model presented by [3]. The base of the location model is to factorize the joint distribution of both outcomes and fit a univariate model to each component of the factorization.

The longitudinal nature renders the data arising from such studies more susceptible to incompleteness. The problem of missingness has been tackled using many approaches, two of them are discussed in depth in this paper: the stochastic EM (SEM) algorithm and the parametric fractional imputation (PFI) algorithm. The stochastic EM (SEM) algorithm is proposed in [4] and developed in [5,6] and [7]. The parametric fractional imputation (PFI) algorithm is introduced in [8,9,10] and [11] in the context of cross-sectional studies, and in [12] and [13], [2] for longitudinal data.

Fewer techniques have been available when the continuous outcome violates normality assumption. In many instances, the assumption of normality is not realistic because of the presence of atypical observations. This problem is more severe if data have missing values [6]. Shen etal. [14] noted that the impact of the distributional form on results would be crucial in the case of incomplete data. Therefore, fitting different distributions to the outcomes, in the case of missing data, is advantageous. It allows us to study the properties of the proposed techniques when the distribution of the data is misspecified. This enables us to conduct a sensitivity analysis of the results to the distributional assumptions of the responses.

A common alternative to the normal distribution is the heavy tailed distributions. Multivariate heavy tailed distributions are proposed and adapted to replace the

traditional multivariate normal distribution [15-23]. A particular heavy tailed distribution is the t distribution that can be expressed as an infinite mixture of normal distributions [24,25]. This idea is first presented in [26] and advocated in longitudinal context by many studies such as [27] and [28]. This distinguished expression is merited of using standard techniques adapted for normal cases. Kenward [29] studies the sensitivity of the selection model under both normality and t distributional assumptions. Gad and Ahmed [6] investigate the effect of the distributional assumptions of the parameters estimates, obtained by the SEM algorithm, in the existence of intermittent missingness assuming both normal and heavy tailed t distribution.

The purpose of this paper is twofold. The first is to extend the joint analysis of mixed discrete and continuous longitudinal data subject to nonignorable missingness assuming non-Gaussian distributions for the continuous outcome; in particular under heavy tailed distributions. To the best of our knowledge, the SEM algorithm and the PFI algorithm have not been applied in the context of mixed discrete and continuous longitudinal outcomes except in [1] and [2] where the normality assumption for the continuous response is imposed. The second is to explore the effect of the distributional assumptions on the SEM and PFI parameters estimates. Therefore, the change of distributional assumption is adopted as a sensitivity tool.

The rest of the paper is organized as follows. Notations and model assumptions are presented in Section 2. The model and proposed methods are highlighted in Section 3. Section 4 is devoted to the application of the SEM algorithm and the PFI algorithm assuming heavy tailed distribution for the continuous outcome. Three simulation studies are presented in Section 5. Application of the proposed techniques on the Interstitial Cystitis Database is discussed in Section 6. Finally, in Section 7, the article ends with few conclusion comments. The Appendix contains some auxiliary tables.

2. Notations and Assumptions

In this article, the same notations of is adopted. The sequences of the continuous and discrete response variables are denoted as y_{ij} and x_{ij}, whilst the sequence z_{ij} is the p-vector of fully observed covariate for the jth measurement from the ith subject at time points $t = (t_1, \dots, t_n)$, for $i = 1, \dots, m$. In vector form, these variables can be expressed as $Y_i = (y_{i1}, \dots y_{in_i})$, $X_i = (x_{i1}, \dots x_{in_i})$ and $Z_i = (z_{i1}, \dots z_{in_i})$ respectively. The number of time points n_i may be common (n) for all individuals.

A sequence of discrete variable x_{ij}^* is assumed to be a realization of independent sample Poisson distribution with parameter λ_{ij}, $j = 1, \dots, n$. Let x_{i0}^* be a Poisson random variable with parameter λ_{i0} that controls the correlation structure of the discrete responses. It is postulated that x_{ij} is a linear combination of the independent Poisson random variables, that is

$$x_{ij} = x_{ij}^* + a x_{i0}^* \qquad (1)$$

where a is a tuning integer parameter that controls the over-dispersion of the data. It can be seen, from Eq. (1), that $\mathrm{E}(x_{ij}) = \lambda_{ij} + a\lambda_{i0}$, $\mathrm{V}(x_{ij}) = \lambda_{ij} + a^2\lambda_{i0}$ and $\mathrm{Cov}(x_{ij}, x_{ik}) = a^2\lambda_{i0}$ for $j \neq k$, which means the larger values of a, the greater the variance of x_{ij}. Note that $\mathrm{Cov}(x_{ij}, x_{ik})$ is always assumed to be positive.

Using log-linear model, the possible relationship between E(X) and E(Y/X) and the matrix of covariates Z can be modelled as

$$\log\big(E(x)\big) = z_{ij}\beta_1$$

While the general linear model is adopted for linking the conditional expectation E(y/x) to the matrix Z.

The conditional distribution of the Y_i given X_i is presumed to depend on conditional mean μ_i and conditional covariance matrix Σ_i. The conditional mean μ_i is presumed to take the form

$$\begin{aligned} \mu_i &= \mathrm{E}(Y_i|X_i) \\ &= \mathrm{E}(Y_i) + \mathrm{cov}(Y_i, X_i)V(X_i)^{-1}\big(X_i - \mathrm{E}(X_i)\big), \end{aligned} \qquad (2)$$

where $\mathrm{cov}(Y_i, X_i) = \mathrm{cov}\big(y_{ij}, x_{ik}\big)_{0 < j,k \le n}$ and

$$V(X_i) = \mathrm{cov}\big(x_{ij}, x_{ik}\big)_{0 < j,k \le n}.$$

To model the covariance between mixed discrete and continuous longitudinal outcomes, It is assumed that

$$\left.\begin{aligned} x_{ij} &= a_x + b_x x_{ij-1} + \epsilon_x, \\ y_{ij} &= a_y + b_y x_{ij-1} + \epsilon_y, \\ b_y &= \gamma_1 b_x + \gamma_2 \end{aligned}\right\} \qquad (3)$$

where a_x, a_y are constants, b_x, b_y are linear effects of x_{ij-1} on x_{ij} and y_{ij} respectively, ϵ_x, ϵ_y are random errors and γ_1, γ_2 are the parameters that control the effect of b_x on b_y.

From Eq. (3) the covariance between any two measurements of the discrete and continuous responses, for the same individual, at any time point $0 < j, k \le n$ can be formulated as:

$$\mathrm{cov}\big(y_{ij}, x_{ik}\big) = \gamma_1 \mathrm{cov}\big(x_{ij}, x_{ik}\big) + \gamma_2 \mathrm{V}\big(x_{ik}\big). \qquad (4)$$

Using the log-linear model the relationship between X_i and the covariate Z_i is modeled whilst the general linear model is employed to model the relationship between the response Y_i and the covariate Z_i, i.e.

$$\log(\lambda_{i0}) = \omega, \log(\lambda_{ij}) = z_{ij}\beta_1, \mathrm{E}(y_{ij}) = z_{ij}\beta_2, \qquad (5)$$

where ω and β_1 are the mean parameters for the discrete response and β_2 is its counterpart for the continuous response. Let Λ and θ denote the vectors of parameters for X_i and Y_i respectively.

Substituting Eq. (5) in Eq. (1) gives

$$\mathrm{E}(x_{ij}) = \lambda_{ij} + a\lambda_{i0} = \exp\big(z_{ij}\beta_1\big) + a\exp(\omega).$$

substituting Eq. (4), Eq. (5) in Eq. (2), it can be shown that

$$\mathrm{E}(y_{ij}|X_i) = z_{ij}\beta_2 + \left[\gamma_1 \mathrm{cov}\left(x_{ij}, x_{ik}\right) + \gamma_2 \mathrm{V}\left(x_{ik}\right)\right]$$
$$\times \left[\mathrm{cov}\left(x_{ij}, x_{ik}\right)\right]^{-1}\left[X_i - \mathrm{E}\left(X_i\right)\right] \quad (6)$$
$$= z_{ij}\beta_2 + \left[\gamma_1 + \gamma_2 \mathrm{V}\left(x_{ik}\right)\mathrm{cov}\left(x_{ij}, x_{ik}\right)^{-1}\right]\left[X_i - \mathrm{E}\left(X_i\right)\right]$$

Yang and Kang [1] shows that from Eq.(6) the conditional expectation $\mathrm{E}(y_{ij}|X_i)$ equals

$$\mathrm{E}(y_{ij}|X_i) = z_{ij}\beta_2 + \gamma_1\left(x_{ij} - a\lambda_{i0} - \lambda_{ij}\right) + \gamma_2 S_i = \varsigma_{ij}\alpha,$$

where $S_i = \frac{1}{n}\sum_{j=1}^n (x_{ij} - a\lambda_{i0} - \lambda_{ij})$, $\varsigma_{ij} = [z_{ij}, (x_{ij} - a\lambda i0 - \lambda ij, Si,$ and $\alpha = \beta 2, \gamma 1, \gamma 2$.

Assuming missingness, X, Y can be classified into two sub-matrices $\{X_o, X_m\}$ and $\{Y_o, Y_m\}$ where X_o, Y_o represent the observed part, while Y_m and X_m denote the missing observations. let u_{ij} and r_{ij} be two binary variables, combined into n_i vectors R_i and U_i, representing the missingness process in X, Y and parameterized by ι and ϕ respectively. r_{ij} takes value 1 if y_{ij} is observed and 0 otherwise. u_{ij} is defined similarly for the discrete variable X. Both r_{ij} and u_{ij} are assumed to follow Bernoulli distribution with a probability $\pi(X_i, Y_i, \phi)$ and $\pi(X_i, Y_i, \iota)$ respectively.

3. Models and Optimization Methods

Based on the selection model of Diggle and Kenward [30], the joint density function of the outcomes X_i, Y_i and response indicators R_i, U_i can be factorized as

$$f\left(X_i, Y_i, R_i, U_i \mid Z_i\right) = f\left(X_i, Y_i \mid Z_i\right)P\left(R_i, U_i \mid X_i, Y_i\right), (7)$$

where

$$f\left(X_i|Z_i\right)$$
$$= \exp\left\{-\sum_{j=0}^n \lambda_{ij}\right\}\prod_{j=1}^n \frac{\lambda_{ij}^{x_{ij}}}{x_{ij}!}\sum_{T=0}^{s_i}\prod_{j=1}^n \frac{x_{ij}!}{(x_{ij}-aT)!T!}\left(\frac{\lambda_{i0}}{\prod_{j=1}^n \lambda_{ij}^a}\right)^T,$$

and s_i is the integer part of $\left[\frac{1}{a}\min_j\left(x_{ij}\right)\right]$.

The probability of missingness for each missingness process can be modeled using logistic models depending on the measurements at the time of missingness d_i; x_{d_i}, y_{d_i} and a set of unknown parameters ϕ, ι, and the logistic models can be schematized as

$$\mathrm{logit}\left\{P\left(y_{id_i}\text{is missing}\right)\right\} = \phi_0 + \phi_1 x_{d_i} + \phi_2 y_{d_i},$$
$$\mathrm{logit}\left\{P\left(x_{id_i}\text{is missing}\right)\right\} = \iota_0 + \iota_1 x_{d_i} + \iota_2 y_{d_i}. \quad (8)$$

The probabilities of missingness depends on the observations at the time of missingness, such a type of missingness is termed as missing not at random [31].

The observed joint distribution function $f(X_o, Y_o, R, U)$ can be obtained as

$$f\left(X_o, Y_o, R, U\right)$$
$$= \iint f\left(X_o, X_m, Y_o, Y_m, R, U\right)dY_m dX_m.$$

The observed data log-likelihood function is proportional to $f(X_o, Y_o, R, U)$, i.e.

$$l\left(\Theta, \tau|X_o, Y_o, R, U\right) \propto f\left(X_o, Y_o, R, U\right),$$

where Θ is the whole set of parameters pertaining to X and Y and $\tau = [\phi, \iota]$. Calculating the Integrations over the missing part is commonly not possible thereby we exploit the ECM algorithm of [32]. The stochastic variants of the ECM algorithm can be used to obviate the integrations entailed in the E-step of the ECM algorithm. However, the key assumptions of this estimation process are

1. The correct specification of the missingness processes.
2. The correct specification of the responses distributions.

Since both outcomes are not fully observed, the correct distributions of the different outcomes are not known. Many studies criticize the selection model with its sensitivity on the distributional assumptions and recommend studying the impact of changing the responses' distributions on the gained results. See [33,34] and the discussion of [30]. Therefore, it is tempted to incorporate a sensitivity analysis when treating the missingness problem. This can be accomplished by fitting different distributions to the responses and investigating the effect of distribution misspecification on the gained parameters.

4. Inference for Non-Normal Models

Yaseen and Gad [2] introduced the SEM algorithm and the PFI algorithm to analyze mixed discrete and continuous (normally distributed) longitudinal data subject to non-random missingness. We extend both algorithms to the heavy tailed distribution family for the continuous response keeping the Poisson assumption for the discrete outcome.

A family of heavy tailed distribution can be expressed as infinite mixtures of normal distributions as

$$f\left(Y_i \mid X_i, q_i\right) = \int f(Y_i \mid X_i, q_i)h(q_i)dq_i, \quad (9)$$

where $f(Y_i|X_i, q_i)$ is normal distribution with mean μ_i and covariance matrix $\frac{\Sigma_i}{q_i}$ and $h(q_i)$ is a known distribution [28]. If $h(q_i)$ has gamma distribution with parameters $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, $f(Y_i|X_i)$ follows multivariate t distribution with mean parameter μ_i, scale matrix Σ_i and degree of freedom ν, i.e.

$$f\left(Y_i \mid X_i\right) = \frac{\Gamma\left(\nu + n_i/2\right)|\Sigma_i|^{-1/2}}{(\pi\nu)^{\frac{n_i}{2}}\Gamma\left(\frac{\nu}{2}\right)\left[1 + d_i^2/\nu\right]^{\nu+n_i/2}},$$

where $d_i^2 = (Y_i - \mu_i)'\Sigma_i^{-1}(Y_i - \mu_i)$.

Adopting $h(q_i)$ as a mixing distribution as in (9), the complete data for the ith subject are $[\,X_{i,o}, X_{i,m}, q_i, Y_{i,o}, Y_{i,m}, R_i, U_i\,]$ whilst $[X_{i,m}\,,\ q_i\,,\ Y_{i,m}]$ represent the missing data. The complete density function can be written as

$$f\left(X_{i,o}, X_{i,m}, q_i, Y_{i,o}, Y_{i,m}, R_i, U_i \mid \Theta, \upsilon, \tau\right)$$
$$= f\left(X_{i,o}, X_{i,m}|\Lambda\right) f\left(Y_{i,o}, Y_{i,m}|X_i, q_i; \theta\right)$$
$$\times h\left(q_i \mid \upsilon\right) P\left(R_i, U_i|X_i, Y_i; \tau\right)$$

and the complete log-likelihood function can be then

$$l\left(\Theta, \upsilon, \tau|X, Y, q, R, U\right)$$
$$= l_1\left(\Lambda|X\right) + l_2\left(\theta|X, Y, q\right) + l_3\left(\upsilon|q\right) + l_4\left(\tau|R, U\right). \quad (10)$$

Maximizing $l(\Theta, \upsilon, \tau|X, Y, q, R, U)$ can be performed by separately maximizing $l_1(\Lambda|X)$, $l_2(\theta|Y, X)$ and $l_4(\tau|R, U)$

if υ is known as $l_3(\upsilon|q)$ has no information about Θ and τ. Otherwise the four components of the right hand side of (10) are required to be maximized. The MLE estimates can be obtained using the ECM algorithm by maximizing the expectation of the complete log-likelihood function given the observed data and current estimates:

$$Q\left(\Theta, \upsilon, \tau|\Theta_t, \upsilon_t, \tau_t\right)$$
$$= E\left\{l\left(\Theta, \upsilon, \tau|X, Y, q, R, U\right)|X_o, Y_o, R, U; \Theta_t, \upsilon_t, \tau_t\right\}$$
$$= \int \left[\begin{matrix} l\left(\Theta, \upsilon, \tau|X, Y, q, R, U\right) \\ \times f\left(X_m, q, Y_m|X_o, Y_o, R, U; \Theta_t, \upsilon_t, \tau_t\right) \end{matrix} \right] dq\ dY_m dX_m$$

and our primary interest is to get the MLE's of Λ, θ, υ and τ.

The SEM algorithm and the PFI algorithm can be adapted to get the parameters estimates by approximating the E-step of the ECM algorithm as follows.

The SEM Algorithm

The SEM algorithm approximates the E-step by iterating the S-step and M-step specified as follows

The S-Step

The S-step involves simulating the missing data $X_{i,m}, q_i, Y_{i,m}$ for the ith subject from their conditional distribution given the observed data and current estimates; $f\left(X_{i,m}, q_i, Y_{i,m}|X_{i,o}, Y_{i,o}, R_i, U_i\right)$. A proper factorization is

$$f\left(X_{i,m}, q_i, Y_{i,m} \mid X_{i,o}, Y_{i,o}, R_i, U_i\right)$$
$$= f\left(X_{i,m}, Y_{i,m}|X_{i,o}, Y_{i,o}, q_i, R_i, U_i\right) g\left(q_i|X_{i,o}, Y_{i,o}\right),$$
$$= f\left(X_{i,m}|X_{i,o}, U_i\right) f\left(Y_{i,m}|X_i, Y_{i,o}, q_i, R_i\right) g\left(q_i|X_{i,o}, Y_{i,o}\right),$$

taking into account the independency between $X_{i,m}$ and $Y_{i,o}, q_i, R_i$ and also between q_i and R_i, U_i.

Since we need to simulate from three conditional distributions, the S-step is classified into three sub-steps, whence sequential sampling technique is conducted.

S1-step

In the first sub step, the missing discrete vector for the ith subject $X_{i,m}$ is imputed at the $(t + 1)^{th}$ iteration from $f\left(X_{i,m}|X_{i,o}, U_i; \Lambda_t\right)$. Details of the simulation procedures

can be found in [2]. Let the simulated vector at the $(t + 1)^{th}$ iteration be denoted as $\tilde{X}_{i,m,t+1}$.

S2-step

The second step entails simulating q_i from $h(q_i|X_{i,o}, \tilde{X}_{i,m,t+1}, Y_{io}; \upsilon_t)$. Markov Chain Monte Carlo techniques are adopted if direct simulation is not applicable. If $h(q_i)$ has gamma distribution with parameters $(\frac{\nu}{2}, \frac{\nu}{2})$, then it is easy to prove that the conditional distribution $h(q_i|X_{i,o}, \tilde{X}_{i,m,t+1}, Y_{io}; \upsilon_t)$ follows also gamma distribution with parameters $(\frac{\nu + n_i}{2}, \frac{\nu + d_i^2}{2})$ [25]. The simulated value for the ith subject is termed as $\tilde{q}_{i,t+1}$.

S3-step

In the last sub-step, $Y_{i,m}$ is simulated from the density $f\left(Y_{i,m}|X_{i,o}, \tilde{X}_{i,m,t+1}, Y_{i,o}, \tilde{q}_{i,t+1}, R_i; \theta_t\right)$ depending on the previous simulated values as presented in [2]. The simulated vector is denoted as $\tilde{Y}_{i,m,t+1}$.

The simulated data in addition to the observed data constitute a Pseudo complete data and the Q-function can be then approximated at the $(t + 1)^{th}$ iteration as

$$Q\left(\Theta, \nu, \tau|\Theta_t, \nu_t, \tau_t\right)$$
$$= \sum_{i=1}^{m} l(\Theta, \nu, \tau|X_{i,o}, \tilde{X}_{i,m,t+1}, Y_{i,o}, \tilde{Y}_{i,m,t+1}, \tilde{q}_{i,t+1}).$$

The M-step

After performing the S-step, the Q-function $Q(\Theta, \nu, \tau|\Theta_t, \nu_t, \tau_t)$ is maximized using conditional maximization procedures as follows.

1. In the first sub-step parameters pertaining to the discrete response are being updated using the generalized linear model as presented in [2].
2. In the second sub-step parameters of the continuous response, $\beta_2, \gamma_1, \gamma_2$, and the covariance parameters are updated conditionally on the current estimates of the parameters pertaining to discrete response. If the conditional distribution $f(Y_i|X_i)$ follows multivariate t distribution and the continuous response is modelled using the general linear model as in (5), then the errors ϵ_i of model (5) follows multivariate normal distribution, i.e.

$$\epsilon_i \mid X_i \sim MVN\left(0, \frac{\Sigma_i}{q_i}\right).$$

According to [25] and [6], if we transform Y_i, Z_i and ϵ_i as

$$Y_i^* = \sqrt{q_i} Y_i, Z_i^* = \sqrt{q_i} Z_i, \epsilon_i^* = \sqrt{q_i} \epsilon_i,$$

then the conditional distribution $f(Y_i^*|X_i)$ follows the traditional multivariate normal distribution with conditional mean μ_i and conditional covariance Σ_i, whence the Jennrich-Scluchter algorithm [35] can be used to maximize θ allowing for different covariance structures pertaining to continuous variable.

3. In the last sub-step, the vector of parameters τ are updated. Since both indicator variables are binary, the iterative reweighted least square method (IRLS) can be employed to find the MLE's.

The PFI Algorithm

The PFI algorithm is executed in the following steps:

1. For each missingness vector; $X_{i,m}, q_i, Y_{i,m}$, K imputed vectors are simulated from the conditional distribution of the missing data given the observed data and current estimates $f(X_{i,m}, q_i, Y_{i,m} | X_{i,o}, Y_{i,o}, R_i, U_i)$. There is no need to change the imputed values except only in two cases. The first case occurs in the first part of the PFI algorithm which corresponds to the burn-in period of the SEM algorithm. The other case is explained in the third step.

2. Using the kth imputed vectors, $\tilde{X}_{i,m}^1, ..., \tilde{X}_{i,m}^K, \tilde{Y}_{i,m}^1, ..., \tilde{Y}_{i,m}^K, \tilde{q}_i^1, ..., \tilde{q}_i^K$ for the vectors of missingness for the ith individual, the kth replicated data are $[\tilde{X}_i^k, \tilde{Y}_i^k, \tilde{q}_i^k] = [(X_{i,o}, \tilde{X}_{i,m}^k), (Y_{i,o}, \tilde{Y}_{i,m}^k), \tilde{q}_i^K]$. Given the kth replicated data $[\tilde{X}_i^k, \tilde{Y}_i^k, \tilde{q}_i^k]$ and the current estimates Θ_t, ν_t, τ_t, a fractional weight $W_{i,t+1}^k$ is calculated at the $(t+1)^{th}$ iteration for each replicated data $[\tilde{X}_i^k, \tilde{Y}_i^k, \tilde{q}_i^K]$ as

$$W_{i,t+1}^k = \frac{f\left(\tilde{X}_i^k, \tilde{Y}_i^k, R_i, U_i, \tilde{q}_i^K; \Theta_t, \nu_t, \tau_t\right)}{f\left(\tilde{X}_i^k, \tilde{Y}_i^k, R_i, U_i, \tilde{q}_i^K; \Theta_0, \nu_t, \tau_0\right)},$$

where Θ_0, ν_0 and τ_0 are the initial estimates.

3. The missing vectors for the ith subject are re-simulated if any of the fractional weights W_i^k exceeds the fraction $K|c$, where c is a fixed constant. This step is essential to avoid unreasonable values in the simulation step and to prevent outliers to dominate the estimation process.

4. The following step is to incorporate the fractional weights in the approximation of the E-step using the weighted mean of the complete density functions calculated based on all the Kth replicated data as follows

$$\hat{Q}_i^1(\Theta | Y_i, X_i) \cong \sum_{k=1}^{K} W_i^k \ln f(Y_{i,o}, X_{i,o}, \tilde{X}_{i,m}^k, \tilde{Y}_{i,m}^k),$$

$$\hat{Q}_i^2(\nu | q_i) \cong \sum_{k=1}^{K} W_i^k \ln h(\tilde{q}_i^K),$$

$$\hat{Q}_i^3(\tau | R_i, U_i)$$
$$\cong \sum_{k=1}^{K} W_i^k \ln f\left(R_i, U_i | X_{i,o}, Y_{i,o}, \tilde{X}_{i,m}^k, \tilde{Y}_{i,m}^k\right),$$

and then

$$\hat{Q}_i(\Theta, \nu, \tau) = \hat{Q}_i^1(\Theta | Y_i, X_i) + \hat{Q}_i^2(\nu | q_i) + \hat{Q}_i^3(\tau | R_i, U_i).$$

5. The algorithm alternates between the previous steps till convergence. The convergence is achieved when the difference between two successive estimates is less than ε; a pre-specified stopping criterion.

5. Simulation Studies

Three simulation studies have been performed. The first simulation study aims to validate the SEM algorithm and the PFI methods under heavy tailed assumption. The second simulation study investigates the effect of misspecification of normal assumption for the continuous outcome on the SEM and PFI estimates. The distribution of the continuous outcome is treated as a heavy tailed variable despite it was originally generated from normal process. The last simulation study reverses the order of misspecification and measures the effect of misspecification of the heavy tailed assumption by analyzing the continuous outcome generated from multivariate t-distribution as multivariate normal variate.

5.1. The First Simulation Study

In this part, the behavior of the SEM algorithm and the PFI algorithm is studied assuming multivariate Poisson distribution for the discrete outcome and multivariate t distribution for the continuous one. Samples of 500 subjects are generated and constructed as bivariate mixed discrete and continuous outcomes. Five covariates have been used. The covariate matrix Z with five independent variables are generated from the normal distribution with a mean of 3 and a variance of 0.25. A description of the study parameters is given in Table 1. The parameters are fixed at $\beta_1 = [-.5, 1, .5]$, $\omega = 2$, $a = 1$, $\beta_2 = [5, 10, 15]$, $\gamma_1 = [.5], \gamma_2 = [1], m = [200], n = [5]$. The degree of freedom ν is set to 5 and is assumed either to be known or estimated from the data. The autoregressive covariance structure is adopted for the continuous longitudinal outcome with $\rho = 0.7$ and $\sigma = 6.0$. A unified logistic model is used to generate the missingness in both outcomes with parameters $\phi = [-17, 0.11, 0.13]$ as in Eq. (8). The burn-in period for the SEM algorithm is chosen with length 250 to all parameters. For the PFI algorithm, the number of replicates K is set to five. We choose c $= 2K$. The first stage length is 10 iterations. Table 2 and Table 3 present the study results.

Table 1. Parameter description

Parameter	description
ω	Parameter of the Poisson variable X_{i0} used to create the correlation of the discrete values
β_1	The mean parameters of the discrete outcome
γ_1 γ_2	The two parameters used to specify the covariance between the discrete and continuous outcomes
β_2	The mean parameters of the continuous outcome
σ ρ	The covariance parameters of the continuous outcome
ϕ	The missingness parameters
ν	The degree of freedom

Table 2. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, degree of freedom ν is unknown

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.890	-0.055	0.012	1.993	-0.003	0.011
	-0.5	-0.476	-0.048	0.002	-0.499	-0.002	0.002
β_1	1	0.984	-0.016	0.001	0.998	-0.002	0.001
	0.5	0.501	0.002	0.001	0.500	0.000	0.001
γ_1	0.5	0.516	0.032	0.003	0.516	0.032	0.003
γ_2	1	0.85	-0.150	0.013	0.864	-0.136	0.012
	5	5.009	0.002	0.714	5.233	0.047	0.638
β_2	10	9.891	-0.011	0.691	9.836	-0.016	0.618
	15	14.989	-0.001	0.648	14.799	-0.013	0.586
σ	6	6.360	0.060	0.435	6.337	0.056	0.736
ρ	0.7	0.652	-0.069	0.006	0.656	-0.063	0.007
	-17	-16.543	-0.027	5.982	-17.046	0.003	7.783
ϕ	0.11	0.118	0.073	0.0004	0.116	0.055	0.014
	0.13	0.122	-0.062	0.001	0.129	-0.008	0.001
ν	5	4.789	-0.042	2.5	4.662	-0.068	4.78

Table 3. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, degree of freedom ν is known

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.894	-0.053	0.010	1.986	-0.007	0.011
	-0.5	-0.469	-0.062	0.002	-0.500	0.000	0.002
β_1	1	0.982	-0.018	0.001	0.994	-0.006	0.001
	0.5	0.495	-0.010	0.001	0.498	-0.004	0.001
γ_1	0.5	0.503	0.006	0.002	0.520	0.040	0.002
γ_2	1	0.828	-0.172	0.011	0.856	-0.144	0.01
	5	5.276	0.055	0.461	5.164	0.033	0.553
β_2	10	9.717	-0.028	0.492	9.682	-0.032	0.565
	15	14.826	-0.012	0.453	14.742	-0.017	0.565
σ	6	5.778	-0.037	0.285	5.713	-0.048	0.873
ρ	0.7	0.630	-0.100	0.005	0.627	-0.104	0.005
	-17	-14.332	-0.157	4.947	-17.239	0.014	8.897
ϕ	0.11	0.121	0.100	0.0005	0.123	0.118	0.001
	0.13	0.100	-0.231	0.0005	0.116	-0.108	0.001

The results highlight that the behavior of the SEM algorithm and the PFI algorithm under heavy tailed assumptions is reasonable. All the parameters have low relative bias (less than 0.20). The performance of the SEM algorithm and the PFI algorithm is alike except for that the relative bias for most of the mean parameters using the PFI algorithm is relatively lower than its counterpart for the SEM algorithm. However, the MSE for the variance parameter σ under the SEM algorithm is lower than the MSE under the PFI algorithm. Overall, the performance of both algorithms is similar whether the degree of freedom is known or not.

5.2. The Second Simulation Study

The aim of this part is to study the sensitivity of the SEM and PFI estimates against misspecification of the distributional assumption. We simulate the continuous outcome assuming multivariate normal distribution but then we treat it as it follows multivariate t distribution and apply the SEM algorithm and the PFI algorithm in the settings of heavy tailed case. The degree of freedom is assumed to be fixed at values 5 and 15. The results are presented in Table 4 and Table 5.

We compare the results of the above tables with the results of the similar simulation study conducted in [2], Table A-1 in the Appendix, where the distribution of the continuous outcome is well specified, i.e. the continuous outcome is generated and analyzed assuming the correct distribution which is multivariate normal. The comparison presents that the MSE for all estimates grows in size or keeps at its value. This note is clearer in the estimates of the parameters pertaining to continuous outcome using the PFI algorithm. For the relative bias, we can classify the parameters to three different groups. The first group includes the parameters estimates that almost have no change in their values. This group includes most of the parameters estimates which are [$\omega, \beta_1, \gamma_1, \gamma_2, \beta_2, \sigma, \rho, \phi_1$] for the SEM and PFI algorithm. The second group contains ϕ_2 that relatively grows in size for both the SEM algorithm and the PFI algorithm. The last group entails only ϕ_0 that has smaller value for the both algorithms. In general, the absolute change in the relative bias is small for the SEM algorithm and the PFI algorithm. There is no much change comparing the results at different values for the degree of freedom.

Table 4. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, degree of freedom ν is fixed at 5, true distribution is multivariate normal distribution

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.883	-0.059	0.012	1.928	-0.036	0.013
β_1	-0.5	-0.473	-0.055	0.002	-0.485	-0.030	0.002
	1	0.983	-0.017	0.001	0.990	-0.010	0.001
	0.5	0.501	0.001	0.001	0.503	0.006	0.001
γ_1	0.5	0.523	0.045	0.003	0.510	0.020	0.003
γ_2	1	0.835	-0.165	0.013	0.857	-0.143	0.014
β_2	5	4.938	-0.012	0.677	4.905	-0.019	0.521
	10	10.011	0.001	0.651	9.956	-0.004	0.684
	15	14.974	-0.002	0.610	14.968	-0.002	0.662
σ	6	6.109	0.018	0.091	5.606	-0.066	0.052
ρ	0.7	0.665	-0.051	0.007	0.695	-0.007	0.006
ϕ	-17	-19.591	0.152	9.991	-19.114	0.124	12.936
	0.11	0.111	0.006	0.001	0.118	0.073	0.001
	0.13	0.156	0.204	0.001	0.149	0.146	0.001

Table 5. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, degree of freedom ν is fixed at 15, true distribution is multivariate normal distribution

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.903	-0.049	0.011	1.936	-0.032	0.013
β_1	-0.5	-0.477	-0.046	0.002	-0.486	-0.028	0.002
	1	0.988	-0.012	0.001	0.996	-0.004	0.001
	0.5	0.498	-0.004	0.001	0.496	-0.008	0.001
γ_1	0.5	0.522	0.045	0.003	0.525	0.050	0.003
γ_2	1	0.842	-0.158	0.011	0.838	-0.162	0.011
β_2	5	4.941	-0.012	0.646	4.992	-0.002	0.613
	10	9.993	-0.001	0.652	10.028	0.003	0.585
	15	14.991	-0.001	0.578	14.869	-0.009	0.545
σ	6	6.141	0.024	0.057	5.974	-0.004	0.044
ρ	0.7	0.659	-0.059	0.005	0.670	-0.043	0.005
ϕ	-17	-19.124	0.125	9.112	-19.199	0.129	11.072
	0.11	0.109	-0.009	0.001	0.115	0.045	0.001
	0.13	0.152	0.171	0.001	0.151	0.162	0.001

5.3. The Third Simulation Study

We reverse the order of distribution misspecification. Data are generated assuming multivariate Poisson distribution for the discrete outcome and multivariate t distribution using the same settings of the above simulation study with degree of freedom equal to 5. The SEM algorithm and the PFI algorithm are applied treating the continuous outcome as normal distributed variable. The results are presented in Table 6.

Table 6. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, true distribution is multivariate t distribution

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.88	-0.06	0.0126	1.99	-0.005	0.0094
β_1	-0.5	-0.47	-0.06	0.0017	-0.50	0	0.0017
	1	0.98	-0.02	0.0012	1.00	0	0.0012
	0.5	0.50	0.00	0.0010	0.50	0	0.0008
γ_1	0.5	0.49	-0.02	0.0031	0.51	0.02	0.0033
γ_2	1	0.82	-0.18	0.0157	0.71	-0.29	0.0143
β_2	5	5.40	0.08	0.6590	5.27	0.05	0.7905
	10	9.58	-0.04	0.6736	9.65	-0.04	0.7844
	15	14.75	-0.02	0.6612	14.81	-0.01	0.7463
σ	6	7.16	0.19	1.0577	7.22	0.20	0.8196
ρ	0.7	0.64	-0.09	0.0112	0.65	-0.07	0.0118
ϕ	-17	-14.01	-0.18	3.3668	-15.17	-0.11	4.4822
	0.11	0.12	0.09	0.0005	0.13	0.18	0.0007
	0.13	0.10	-0.23	0.0003	0.10	-0.23	0.0004

Results from this study and results from Table 2 are brought together and contrasted. Some notes can be concluded. The relative bias for the parameters under misspecification of multivariate t distribution hardly differs from its counterpart under well-specification of multivariate t distribution. The exception for that is the variance parameter σ where its absolute relative bias jumps form .03, .04 to .19, .20 for the SEM and PFI algorithms respectively. For the MSE values, the estimates under misspecification of the multivariate t distribution are higher or the same as the MSE values compared to well-specification case.

The main conclusion of the sensitivity simulation studies is that misspecification of the distribution of the continuous outcome does not affect the relative bias of the estimates- except for the variance parameter. However, the MSE grows in size which leads into a loss of significance for truly significant variables.

6. Application (Interstitial Cystitis Database)

The proposed methods are applied under the assumption of heavy tailed distribution to the Interstitial Cystitis Data Base (ICDB) and the results are compared with the similar results calculated under normality assumptions presented in [2] which is presented in Table A-2 in the Appendix. The ICDB data have been used by [1]. The ICDB characteristics are discussed in details in [36]. The ICDB data include 637 patient at its baseline. Patients are followed for symptoms of pain, urgency and urinary frequency, from January 1993 to November 1997. Yang and Kang [1] study the joint effect of a group of covariates on the urgency and urinary frequency treating them as continuous and discrete variables respectively. Each of these variables are measured by asking the patients to rate them in the last week on an ordinal scale ranging from 0; for the lowest severity, to 9 which is the maximum severity. In addition, the patients are required to rate the same variables in three consecutive days. The averages of the study variables over the three days are also recorded. The main problem of the data is the high rate of missingness. About 20% of the sample only remained in the study after the 36th month. Therefore, [1] consider only the data gathered in the first 36 months.

In this article we use the same settings of [2] but assuming multivariate heavy tailed distribution for the continuous outcome. Patients with completely missing data or data with outliers in any of the outcomes are omitted from the analysis leading to a reduced sample of 538 patients with maximum observed time points equals 44. A brief description of the covariates is given in Table A-3 in the Appendix.

The analysis starts with similar settings for the first simulation study. The number of replicates for the PFI is set to 3. The standard errors of the estimates are obtained using the Jackknife replication method and the p-values are produced based on the Z-table. The same burn-in period as in the simulation studies are used for the SEM algorithm. The results are presented in Table 7.

For the SEM algorithm, a comparison of the results under normal assumption and heavy tailed assumption shows no difference in terms of the relative bias. For the significance levels, only Shx_2 decreases in its significance from being significant at levels .01 to being significant at level 0.1. Other variables attain its significance at approximately the same levels. For the PFI estimates, the same relative biases are found between the results under different distributional assumptions. However, many variables lose their significance under heavy tailed assumption compared to normal estimates. This includes income, Shx_2, UROD_7, sev_2, in β_1 and sex, age in β_2. The general conclusion from both results is that assuming heavy tailed distribution instead of normal distribution leads to an increase in the standard errors thereby many variables lose their significance. However, the SEM algorithm is more robust than the PFI algorithm since the change in the significance levels is limited which is a merit over the PFI algorithm. These conclusions coincide with the results of the simulation studies where the change in the distributional assumption affects only the MSE not the relative bias. Other interesting note is the type of missingness the data suffer from. Under multivariate normal assumption, the missingess is considered MNAR according to both SEM and PFI algorithm. In contrast, the nonignorable missingess parameters are insignificant under multivariate t distribution which is also a common result between our study and [6].

Table 7. parameter estimates of the ICDB data and its significance

Parameter	Variable	PFI	SEM
ω		1.35	1.26***
B_1	sex	0.33***	0.32***
	income	0.17	0.20***
	Shx_2	0.09	0.11*
	UROD_7	-0.0008	-0.0008**
	UROD_9	-0.0002	-0.0001
	age	0.01*	0.01***
	sev_2	0.75	0.74***
	sev_3	1.04*	1.00***
γ_1		0.15	0.15
γ_2		-0.003	-0.004
B_2	sex	0.84	0.79***
	income	-0.21	-0.23
	shx_2	-0.02	-0.03
	UROD_7	-0.0013	-0.001
	UROD_9	0.0004	0.0002
	age	0.01	0.01***
	sev_2	2.11***	2.19***
	sev_3	3.26***	3.37***
ϕ_0		-2.54	-2.14**
ϕ_1		0.03	0.01
ϕ_2		0.08	0.07
σ		1.70	1.65
ρ		.39	0.40

*** significant at 0.01 significance level, ** significant at 0.05 significance level, *significant at 0.10 significance level.

7. Concluding Remarks

In this article, the SEM algorithm and the PFI algorithm are extended to heavy tailed case for the continuous outcome. The proposed methods are advantageous in the sense of low relative bias and MSE. The Jennrich-Schluchter algorithm incorporated in the estimation of the covariance parameters permits more flexibility in the specification of the covariance matrix related to the continuous outcome. The sensitivity analysis shows that misspecification of the outcome's distribution does not affect the values of the estimates. Parameters estimates are still unbiased but their standard errors are increased. This may affect the significance of the variables especially with the PFI algorithm. Yet there are more possible extensions need further investigation. The sensitivity analysis should be investigated against the distributional assumptions for the discrete outcome as well. This is applicable for instance by adopting multivariate negative binominal distribution instead of the multivariate Poisson distribution for the discrete outcome. Validating different mechanisms and forms for the assumed missingness mechanism is also advisable. Another important further point is to represent the covariance of the discrete outcome using a matrix of parameters instead of one as discussed by [37] and [38].

Important Note

The ICDB data reported here were supplied by the NIDDK Central Repositories. This manuscript does not necessarily reflect the opinions or views of the NIDDK Central Repositories, or the NIDDK.

References

[1]   Yang, Y. and J. Kang, Joint analysis of mixed Poisson and continuous longitudinal data with nonignorable missing values. Computational Statistics & Data Analysis, 2010. 54(1): p. 193-207.

[2]   Yaseen, A.S.A. and A.M. Gad, A stochastic variant of the EM algorithm to fit mixed (discrete and continuous) longitudinal data with nonignorable missingness. Communications in Statistics - Theory and Methods, 2019.

[3]   Olkin, I. and R.F. Tate, Multivariate correlation models with mixed discrete and continuous variables. The Annals of Mathematical Statistics, 1961: p. 448-465.

[4]   Celeux, G. and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. Computational statistics quarterly, 1985. 2(1): p. 73-82.

[5]   Gad, A.M. and A.S. Ahmed, Analysis of longitudinal data with intermittent missing values using the stochastic EM algorithm. Computational Statistics & Data Analysis, 2006. 50(10): p. 2702-2714.

[6]   Gad, A.M. and A.S. Ahmed, Sensitivity analysis of longitudinal data with intermittent missing values. Statistical Methodology, 2007. 4(2): p. 217-226.

[7]   Gad, A.M. and N.I. EL-Zayat, Fitting Multivariate Linear Mixed Model for Multiple Outcomes Longitudinal Data with Non-ignorable Dropout. International Journal of Probability and Statistics, 2018. 7(4): p. 97-105.

[8]   Kim, J.K. and W. Fuller, Parametric fractional imputation for missing data analysis. Joint Statistical Meeting Proceedings, 2008: p. 158-169.

[9]   Kim, J.K., Parametric fractional imputation for missing data analysis. Biometrika, 2011. 98(1): p. 119-132.

[10] Kim, J.Y. and J.K. Kim, Parametric fractional imputation for nonignorable missing data. Journal of the Korean Statistical Society, 2012. 41(3): p. 291-303.

[11] Kim, J.K. and M. Hong, Imputation for statistical inference with coarse data. Canadian Journal of Statistics, 2012. 40(3): p. 604-618.

[12] Yang, S., J.-K. Kim, and Z. Zhu, Parametric fractional imputation for mixed models with nonignorable missing data. Statistics and Its Interface, 2013. 6(3): p. 339-347.

[13] Yaseen, A. S., Gad, A. M., & Ahmed, A. S, Maximum Likelihood Approach for Longitudinal Models with Nonignorable Missing Data Mechanism Using Fractional Imputation. American Journal of Applied Mathematics and Statistics, 2016. 4(3): p. 59-66.

[14] Shen, S., C. Beunckens, C. Mallinckrodt, and G. Molenberghs, A local influence sensitivity analysis for incomplete longitudinal depression data. Journal of biopharmaceutical statistics, 2006. 16(3): p. 365-384.

[15] Pinheiro, J.C., C. Liu, and Y.N. Wu, Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. Journal of Computational and Graphical Statistics, 2001. 10(2): p. 249-276.

[16] Wang, W.-L. and T.-H. Fan, Estimation in multivariate t linear mixed models for multiple longitudinal data. Statistica Sinica, 2011: p. 1857-1880.

[17] Luo, S., J. Ma, and K.D. Kieburtz, Robust Bayesian inference for multivariate longitudinal data by using normal/independent distributions. Statistics in medicine, 2013. 32(22): p. 3812-3828.

[18] Wang, W.-L., T.-I. Lin, and V.H. Lachos, Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. Statistical methods in medical research, 2018. 27(1): p. 48-64.

[19] Achcar, J.A., E.A. Coelho-Barros, J.R.T. Cuevas, and J. Mazucheli, Use of Lèvy distribution to analyze longitudinal data with asymmetric distribution and presence of left censored data. Communications for Statistical Applications and Methods, 2018. 25(1): p. 43-60.

[20] Lee, D., Y. Lee, M.C. Paik, and M.G. Kenward, Robust inference using hierarchical likelihood approach for heavy-tailed longitudinal outcomes with missing data: An alternative to inverse probability weighted generalized estimating equations. Computational statistics & data analysis, 2013. 59: p. 171-179.

[21] Wang, W.-L., Mixture of multivariate t nonlinear mixed models for multiple longitudinal data with heterogeneity and missing values. TEST, 2018. 28: p. 1-27.

[22] Wang, W.-L. and T.-H. Fan, Bayesian analysis of multivariate t linear mixed models using a combination of IBF and Gibbs samplers. Journal of Multivariate Analysis, 2012. 105(1): p. 300-310.

[23] Wang, W.L. and T.I. Lin, Multivariate t nonlinear mixed‐effects models for multi-outcome longitudinal data with missing values. Statistics in medicine, 2014. 33(17): p. 3029-3046.

[24] Peel, D. and G.J. McLachlan, Robust mixture modelling using the t distribution. Statistics and computing, 2000. 10(4): p. 339-348.

[25] McLachlan, G. and T. Krishnan, The EM algorithm and extensions. Vol. 382. 2007: John Wiley & Sons.

[26] Andrews, D.F. and C.L. Mallows, Scale mixtures of normal distributions. Journal of the Royal Statistical Society. Series B (Methodological), 1974: p. 99-102.

[27] Meza, C., F. Osorio, and R. De la Cruz, Estimation in nonlinear mixed-effects models using heavy-tailed distributions. Statistics and Computing, 2012. 22(1): p. 121-139.

[28] Forbes, F. and D. Wraith, A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. Statistics and Computing, 2014. 24(6): p. 971-984.

[29] Kenward, M.G., Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. Statistics in medicine, 1998. 17(23): p. 2723-2732.

[30] Diggle, P. and M.G. Kenward, Informative drop-out in longitudinal data analysis. Applied statistics, 1994: p. 49-93.

[31] Little, R.J. and D.B. Rubin, statistical analysis with missing data. 1987, New York: Wiley.

[32] Meng, X.-L. and D.B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 1993. 80(2): p. 267-278.

[33] Council, N.R., Principles and methods of sensitivity analyses, in The Prevention and Treatment of Missing Data in Clinical Trials. 2010, National Academies Press (US).

[34] Daniels, M.J., D. Jackson, W. Feng, and I.R. White, Pattern mixture models for the analysis of repeated attempt designs. Biometrics, 2015. 71(4): p. 1160-1167.

[35] Jennrich, R.I. and M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices. Biometrics, 1986. 42: p. 805-820.

[36] Propert, K.J., A.J. Schaeffer, C.M. Brensinger, J.W. Kusek, L.M. Nyberg, and J.R. Landis, A prospective study of interstitial cystitis: results of longitudinal followup of the interstitial cystitis data base cohort. The Journal of urology, 2000. 163(5): p. 1434-1439.

[37] Karlis, D. and L. Meligkotsidou, Multivariate Poisson regression with covariance structure. Statistics and Computing, 2005. 15(4): p. 255-265.

[38] Shi, P. and E.A. Valdez, Multivariate negative binomial models for insurance claim counts. Insurance: Mathematics and Economics, 2014. 55: p. 18-29.

Appendex

Table A-1. Parameter estimates, relative bias and mean square error of the SEM and PFI estimates, Yaseen and Gad (2019) results

Parameter	True Value	SEM Estimates			PFI Estimates		
		Estimate	Relative Bias	MSE	Estimate	Relative Bias	MSE
ω	2	1.86	0.07	0.011	1.97	0.01	0.012
β_1	-0.5	-0.47	-0.06	0.002	-0.49	-0.01	0.002
	1	0.98	0.02	0.001	1.00	0.00	0.001
	0.5	0.51	0.02	0.001	0.50	0.01	0.001
γ_1	0.5	0.5	0.00	0.002	0.49	0.01	0.002
γ_2	1	0.82	0.18	0.010	0.84	0.16	0.010
β_2	5	5.17	0.03	0.603	5.37	0.07	0.539
	10	9.71	0.03	0.359	9.66	0.03	0.454
	15	14.94	0.00	0.561	14.71	0.02	0.480
σ	6	5.82	0.03	0.022	5.81	0.03	0.042
ρ	0.7	0.63	0.10	0.005	0.62	0.12	0.005
ϕ	-17	-15.02	-0.12	2.984	-14.53	-0.15	3.773
	0.11	0.12	0.09	0.001	0.12	0.06	0.001
	0.13	0.11	0.15	0.0003	0.10	0.21	0.0004

Table A-2. Parameter estimates of the ICDB data and its significance under normality assumption, Yaseen and Gad (2019) results

Parameter	Variable	PFI (compound symmetric)	PFI (autoregressive)	SEM (compound symmetric)	SEM (autoregressive)
ω		1.34***	1.34***	1.26***	1.26***
B_1	sex	0.30***	0.35***	0.32***	0.32***
	income	0.20**	0.18**	0.20***	0.20***
	Shx_2	0.12**	0.10**	0.10***	0.10***
	UROD_7	-0.001**	-0.001**	-0.001**	-0.001**
	UROD_9	-0.0002	-0.0002	-0.0001	-0.0001
	age	0.01***	0.01***	0.01***	0.01***
	sev_2	0.73***	0.69***	0.73***	0.73***
	sev_3	1.00***	0.97***	0.99***	1.00***
γ_1		0.18	0.16	0.16	0.18
γ_2		-0.05	-0.04	-0.04	-0.05
B_2	sex	0.78***	0.79***	0.79***	0.80***
	income	-0.08	-0.07	-0.07	-0.08
	shx_2	0.11	0.06	0.03	0.04
	UROD_7	-0.001	-0.001	-0.001	-0.001
	UROD_9	0.0004	0.0003	0.0001	0.0003
	age	0.01***	0.01***	0.01*	0.01*
	sev_2	2.11***	2.09***	2.16***	2.17***
	sev_3	3.21***	3.14***	3.20***	3.26***
σ		1.55	1.81	1.80	1.53
ρ		.50	1.78	0.49	1.78
ϕ_0		-2.65***	-2.52***	-2.16***	-2.25***
ϕ_1		0.04*	0.04*	0.01	0.004
ϕ_2		0.10*	0.05*	0.07***	0.10***

*** significant at 0.01 significance level, ** significant at 0.05 significance level, *significant at 0.10 significance level.

Table A-3. The definition of the covariates used in ICDB data

variable	definition	type
Sex	Sex: male/ female	binary
income	Annual household income	continuous
shx_2	Previous interstitial cystitis Binary diagnosis by physician: yes/no	binary
UROD_7	Volume at first sensation	continuous
UROD_9	Volume at maximal capacity	continuous
age	Patient age	continuous
Severity	Severity of symptoms with three levels: sev_1, sev_2, sev_3	Categorical: three levels