Published in final edited form as:

Stat Med. 2015 July 20; 34(16): 2391–2402. doi:10.1002/sim.6514.

A Random Pattern Mixture Model for Ordinal Outcomes with Informative Dropouts

Chengcheng Liu^{1,*}, Sarah J. Ratcliffe², and Wensheng Guo²

¹Allergan, Inc. 200 Somerset Corporate Blvd, Suite 6001, Bridgewater, NJ 08807, U.S.A

²Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104-6021, U.S.A.

Abstract

We extend a random pattern mixture joint model for longitudinal ordinal outcomes and informative dropouts. The patients are generalized to "pattern" groups based on known covariates that are potential surrogates for the severity of the underlying condition. The random pattern effects are defined as the latent effects linking the dropout process and the ordinal longitudinal outcome. Conditional on the random pattern effects, the longitudinal outcome and the dropout times are assumed independent. Estimates are obtained via the EM algorithm. We applied the model to the end-stage renal disease (ESRD) data. Anemia was found to be significantly affected by baseline iron treatment when the dropout information was adjusted via the study model; as opposed to an independent or shared parameter model. Simulations were performed to evaluate the performance of the random pattern mixture model under various assumptions.

Keywords

Adaptive gaussian quadrature; EM algorithm; Newton-Raphson algorithm; Pattern mixture model; Ordinal outcome

1. Introduction

Ordinal responses are very important in biomedical studies. Subjects can be classified into groups based on the severity of a disease or condition. For example, depression is often classified as none, mild, moderate or severe. These groupings are generally more clinically relevant than underlying continuous scores. Random effect ordinal regression models have been widely used when the observations are not independent [1, 2]. In longitudinal studies, ordinal response models can suffer from the same bias issues as continuous outcomes when informative dropouts exist and the dropout mechanism is ignored. Guo and others first introduced a random pattern mixture model for gaussian longitudinal data with informative dropouts [3]. Unlike traditional pattern mixture model, the patterns in the random pattern mixture model do not have to be based on dropout time. Instead, the patterns are defined based on observed covariates that are considered to be a surrogate for disease severity. The

* Correspondence to: Allergan, Inc. 200 Somerset Corporate Blvd, Suite 6001, Bridgewater, NJ 08807, U.S.A.
liuch@mail.med.upenn.edu.

random pattern variable is then a latent variable linking the outcome and the time to dropout, and by stratifying the data into patterns the outcome and the time to dropout are conditionally independent. In this paper, we apply the random pattern mixture model to ordinal longitudinal responses with informative dropouts.

Joint modeling has been widely used to correct biases when handling nonignorable missing data (NMAR) [4, 5, 6, 7, 8, 9, 3]. Most of the joint models were developed for normally distributed longitudinal data, though several papers have been published on joint models with respect to non-gaussian data [10, 11, 12, 13, 14, 15, 16]. Two research groups have extended the shared parameter model on longitudinal binary to ordinal outcomes. In extension of the work on binary outcomes [11], TenHave and others proposed a shared parameter model with a mixed effects ordinal logistic model for the longitudinal ordinal response data and a continuation ratio logit (CRL) models for the dropout [17]. Pulkstenis and others extended the shared parameter longitudinal binary model [12] to an ordinal response [18]. A closed form marginal likelihood was obtained by using a log-log link for both ordinal response and the binary survival data as well as a log-gamma distribution of the random effects.

The study random pattern mixture model for ordinal longitudinal responses extends the pattern-mixture framework to the probit and logit link functions. This model has several advantages over existing models. Firstly, by linking the ordinal response and dropout times at the pattern level more flexibility and efficiency is introduced. The patterns can be generated using clinically meaningful data that helps divide the population into groups base on underlying disease severity. Information can then be "borrowed" between subjects within the same pattern, gaining more efficient estimates. Additionally, this "borrowing" allows subjects with only a very small number of longitudinal observations to be included in the estimation. Models linked at the subject level often require a minimum of 3, or more, longitudinal observation per subject in order to find stable estimates[19], whereas the random pattern mixture model can incorporate any number of observations per subject . The random pattern mixture model gains efficiency by grouping multiple subjects into a pattern. When there are very few subjects per pattern, the random pattern mixture model approaches the shared parameter model and the improved efficiency may be limited. Next, both the standard subject level shared parameter model and the pattern mixture model are special cases of the study model. If each subject becomes their own pattern, then study model reduces to the more traditional shared parameter model. The traditional pattern mixture model can be estimated using the study model with patterns defined by the dropout times. However, the random pattern mixture model finds the final estimate as the integral over the random pattern effects and therefore avoids the over-parameterization problem inherent in traditional pattern mixture models even when dropout times are used to define the patterns. Finally, traditional subject level selection (or pattern mixture) models implicitly impute the longitudinal outcomes after dropout in order to prevent biased estimates. This assumes the individual would have continued on the same underlying disease trajectory if they would still be observed. In the case of dropout due to death, this can be conceptually problematic. Subjects who died often do so once they pass a maximum disease threshold and interpreting treatment levels after death do not make sense. By linking at the pattern level, this conceptual issue is removed. The pattern-level link parameters can now be thought of the

increase in the outcome/death rate for subjects who start with a certain level of underlying disease, of which the pattern is a surrogate.

Our research was motivated by the end-stage renal disease (ESRD) data [20]. Erythropoietin has been used for ESRD patients to treat anemia. However many patients have been found to develop iron deficiency during the initial phase of administration of erythropoietin. Parental iron is often used to increase levels of hemoglobin and reduce the use of erythropoietin. The objective of the analysis was to examine the short and longer term effect of baseline iron dosing on the severity of anemia. Approximately one-third of the analytic sample died during the course of the study. Ignoring, or incorrectly modeling, the dropout mechanism in this study could cause biased estimates. Additionally, with some subjects having limited follow-up, the "borrowing" across subjects within a pattern of the pattern mixture model is a needed feature. A random pattern mixture model with age as a pattern variable and a shared parameter model are compared in the application.

Section 2 and Section 3 describe the random pattern mixture model with ordinal outcome and estimation method. Application to a dataset from ESRD data and simulations are presented in Section 4 and Section 5. Conclusion remarks are provided in Section 6.

2. A Random Pattern Mixture Model with Ordinal Outcome

In the random pattern mixture model, the longitudinal outcome and dropout time were assumed independent conditional on the random pattern effects. Here we apply the random pattern mixture model to ordinal outcomes. Data are stratified into N patterns. Within the i -th pattern there are n_i subjects, and the j -th subject has n_{ij} measurements. Let y_{ijk} be the S level ordinal outcome for the k -th measurement of the j -th subject within the i -th pattern. Then, $\mathbf{y}_{ijk} = [y_{ijk1}, y_{ijk2}, \dots, y_{ijkS}]^T$, with each value 0 or 1. Let $\pi_{ijk} = [\pi_{ijk1}, \dots, \pi_{ijkS}]^T$ be the multinomial probabilities corresponding to $\mathbf{y}_{ijk}$. Let λ_{ijkI} be the I -th cumulative probability

with $\lambda_{ijkI} = \sum_{l'=1}^I \pi_{ijk l'}$, and $I = 1, \dots, S-1$. We model the longitudinal ordinal response using a proportional odds model with random subject, β_{ij} , and pattern, $\mathbf{u}_i$, effects.

Let r_{ij} be the time to dropout for the j -th subject within the i -th pattern. We model the dropout times via a linear mixed effects model with pattern level random effects. These pattern level random effects are shared with the ordinal longitudinal outcome, linking the two models at the pattern level. Thus, the final model is:

$$\begin{aligned} \mathbf{y}_{ijk} | \pi_{ijk} &\sim \text{Multinomial}(\pi_{ijk}) \\ \text{logit}(\lambda_{ijkI}) &= \alpha_{0I} + \mathbf{x}_{ijk}^T \boldsymbol{\alpha}_1 + \mathbf{z}_{ijk}^T \boldsymbol{\beta}_{ij} + \mathbf{w}_{ijk}^T \mathbf{u}_i \\ \pi_{ijkS} &= 1 - \sum_{l=1}^{S-1} \pi_{ijk l} \\ r_{ij} &= \mathbf{x}_{2ij}^T \boldsymbol{\alpha}_2 + \mathbf{b}^T \mathbf{u}_i + \varepsilon_{ij} \end{aligned}$$

where

$$\begin{aligned}\beta_{ij} &\sim N(0, \Sigma_\beta), \\ \mathbf{u}_i &\sim N(0, \Sigma_u), \\ \varepsilon_{ij} &\sim N(0, s^2),\end{aligned}$$

and a_{0l} is the intercept for the l -th cumulative probability logistic regression; $0 < a_{01} < \dots < a_{0S}$. To insure a_{0l} are positive, we instead estimate the exponential form of ρ_{0l} , $a_{0l} = \exp(\rho_{0l})$. The known design matrices are: $\mathbf{x}_{1ijk}$, associated with the longitudinal fixed effects α_1 ; $\mathbf{z}_{ijk}$ and $\mathbf{w}_{ijk}$ for the unknown subject and pattern level random effects; and $\mathbf{x}_{2ij}$ associated with the dropout time fixed effects α_2 . Finally, ε_{ij} is the residual for r_{ij} and $\mathbf{b}$ is an unknown parameter vector linking $\mathbf{u}_i$ to r_{ij} .

For simplicity, we have assumed that the intercept for each level of the ordinal outcome is equal across patterns. In many situations this assumption is not too restrictive since differences by pattern can be handled by inclusion of pattern-ordinal level combinations in the fixed effects design matrix and/or the pattern level random effects. When the assumption is restrictive, the model can be easily adapted to a vector for a_{0l} such that $a_{0l} = [a_{0il}]^T$ and a design matrix added. Additionally, this extension results in the model not being restricted to the proportionality of the odds assumption. The proportional odds model assumes the common slope β_{ij} across levels. When there is no common slope across levels, the assumption is violated. To accommodate the departure of the assumption, an interaction can be added which allows different β_{ijl} at the l -th level. The model can be adapted by $\beta_{ij} = [\beta_{ijl}]^T$ and adding level indicator in the design matrix $\mathbf{z}_{ijk}$.

Let $\mathbf{v} = (\rho_{01}, \dots, \rho_{0(S-1)}, \alpha_1)$. Then, the set of parameters to be estimated in this model is $\phi = (\mathbf{v}, \alpha_2, \mathbf{b}, \Sigma_\beta, \Sigma_u, s^2)$. A shared parameter model is a special case of the random pattern mixture model when each subject is its own pattern. If the patterns are defined based upon the time to dropout, then the model is equivalent to a pattern mixture model where the pattern effects are integrated out rather than averaged.

The dropout times are often log-transformed in order to meet the normality assumption. We will use r_{ij} throughout the paper for notational simplicity, even if $\log(r_{ij})$ is being estimated. Further, study follow-up is generally completed before the dropout time for all subjects is observed. In this case, we observe the censor time, c_{ij} , rather than r_{ij} . Thus, the observed dropout time is $r_{ij}^* = \min(r_{ij}, c_{ij})$. The dropout indicator, δ_{ij} , differentiates between observed dropout times and censoring time.

3. Estimation

The random pattern mixture model is estimated via the EM algorithm [21] with a nested Newton-Raphson algorithm. The maximum likelihood estimates (MLEs) are calculated by treating $\mathbf{u}_i$, β_{ij} and censored r_{ij} as missing data. The expectation of the complete data log likelihood is calculated with respect to $\mathbf{u}_i$, β_{ij} and censored r_{ij} given the observed data $(\mathbf{y}_{ij}, r_{ij}^*, \delta_{ij})$ and parameter estimates $\hat{\phi}$ in the E step and then maximized in the M step.

The complete data log likelihood is given by

$$l_c = \sum_i^N \sum_j^{n_i} \log f(y_{ij} | \mathbf{X}_{1ij}, \mathbf{u}_i, \boldsymbol{\beta}_{ij}, \hat{\boldsymbol{\varphi}}) f(r_{ij} | \mathbf{x}_{2ij}, \mathbf{u}_i, \hat{\boldsymbol{\varphi}}) \times f(\mathbf{u}_i | \hat{\boldsymbol{\varphi}}) f(\boldsymbol{\beta}_{ij} | \hat{\boldsymbol{\varphi}})$$

where

$$f(y_{ij} | \mathbf{X}_{1ij}, \mathbf{u}_i, \boldsymbol{\beta}_{ij}, \hat{\boldsymbol{\varphi}}) = \prod_k^{n_{ij}} \left[\left(\frac{\exp(\alpha_{01} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{01} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} \right)^{y_{ijk1}} \times \prod_{l=2}^{S-1} \left(\frac{\exp(\alpha_{0l} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{0l} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} - \frac{\exp(\alpha_{0(l-1)} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{0(l-1)} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} \right)^{y_{ijk}} \right]$$

$$G_{ijk} = \mathbf{z}_{ijk}^T \boldsymbol{\beta}_{ij} + \mathbf{w}_{ijk}^T \mathbf{u}_i$$

and $f(\mathbf{u}_i | \cdot)$, $f(\boldsymbol{\beta}_{ij} | \cdot)$ and $f(r_{ij} | \cdot)$ follow gaussian density functions. Thus the complete data log likelihood is proportional to

$$l_c \propto \sum_i^N \sum_j^{n_i} \log \left[\prod_k^{n_{ij}} \left[\left(\frac{\exp(\alpha_{01} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{01} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} \right)^{y_{ijk1}} \times \prod_{l=2}^{S-1} \left(\frac{\exp(\alpha_{0l} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{0l} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} - \frac{\exp(\alpha_{0(l-1)} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})}{1 + \exp(\alpha_{0(l-1)} + \mathbf{x}_{1ijk}^T \boldsymbol{\alpha}_1 + G_{ijk})} \right)^{y_{ijk}} \right] \right]$$

$$- \frac{1}{2} \sum_i^N \sum_j^{n_i} \left[\log s^2 + \left(r_{ij} - \mathbf{x}_{2ij}^T \boldsymbol{\alpha}_2 - \mathbf{b}^T \mathbf{u}_i \right)^T s^{-2} \left(r_{ij} - \mathbf{x}_{2ij}^T \boldsymbol{\alpha}_2 - \mathbf{b}^T \mathbf{u}_i \right) \right]$$

$$- \sum_i^N \frac{n_i}{2} \left[\log |\boldsymbol{\Sigma}_u| + \mathbf{u}_i^T \boldsymbol{\Sigma}_u^{-1} \mathbf{u}_i \right]$$

$$- \frac{1}{2} \sum_i^N \sum_j^{n_i} \left[\log |\boldsymbol{\Sigma}_\beta| + \boldsymbol{\beta}_{ij}^T \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}_{ij} \right]$$

In the E step, the expectation of the complete data log likelihood is taken. The resulting sufficient statistics to be calculated are:

$$\sum_i \mathbf{X}_{2i} \mathbf{u}_i, \sum_i \mathbf{X}_{2i}^T \mathbf{r}_i \sum_i \mathbf{r}_i \mathbf{u}_i, \sum_i n_i \mathbf{u}_i \mathbf{u}_i^T, \\ \sum_{ij} \boldsymbol{\beta}_{ij} \boldsymbol{\beta}_{ij}^T, \sum_i \mathbf{r}_i^T \mathbf{r}_i, \sum_i \mathbf{r}_i, \sum_i n_i \mathbf{u}_i$$

where $\mathbf{X}_{2i}$ is the design matrix combining $\mathbf{x}_{2ij}$ in the i -th pattern, and $\mathbf{r}_i$ is the vector of r_{ij} in the i -th pattern. The expected sufficient statistics are expectations of

$E(\cdot | \mathbf{X}_{1ij}, \mathbf{x}_{2ij}, r_{ij}, \delta_{ij}, \hat{\boldsymbol{\varphi}})$, though we will write $E(\cdot)$ for notational simplicity throughout.

Closed form expressions for the expectation of the sufficient statistics conditional on $\hat{\boldsymbol{\varphi}}$ are not possible. Hence, the expectations are calculated using adaptive gaussian quadrature [22] to perform the necessary integrations. Expectations with respect to $\mathbf{r}_i$ are only required for subjects who are censored. In this case, $f(r_{ij} | r_{ij} > c_{ij}, \cdot)$ is a truncated normal density and properties of this distribution can be used to estimate the expectation and avoid the integration over censored r_{ij} . The expectations then become two level integrals; subject level $\boldsymbol{\beta}_{ij}$ first and then pattern level $\mathbf{u}_i$. Adaptive gaussian quadrature centers the grid of abscissas around the conditional mode and uses the conditional standard unit deviation as the scale of

the abscissas. Conditional modes and standard deviations are obtained by maximizing the complete distribution of the observed and missing data.

The expected sufficient statistics are then used to calculate the estimates in the M step. The Newton-Raphson algorithm is used to calculate $\boldsymbol{\nu}$, all others have closed form solutions.

$$\begin{aligned}\boldsymbol{\nu}^{(k+1)} &= \boldsymbol{\nu}^{(k)} + \left(-\frac{\partial^2 E(l_c)^{(k)}}{\partial \boldsymbol{\nu}^2} \right)^{-1} \frac{\partial E(l_c)^{(k)}}{\partial \boldsymbol{\nu}} \\ \hat{\boldsymbol{\alpha}}_2 &= \left(\sum_i \mathbf{X}_{2i}^T \mathbf{X}_{2i} \right)^{-1} E \left(\sum_i \mathbf{X}_{2i}^T (\mathbf{r}_i - \mathbf{1}_{n_i} \mathbf{b}^T \mathbf{u}_i) \right) \\ \hat{\mathbf{b}} &= E \left(\sum_i n_i \mathbf{u}_i \mathbf{u}_i^T \right)^{-1} E \left(\sum_i \mathbf{u}_i \mathbf{1}_{n_i}^T (\mathbf{r}_i - \mathbf{X}_{2i} \boldsymbol{\alpha}_2) \right) \\ \hat{\boldsymbol{\Sigma}}_u &= \left(\sum_i n_i \right)^{-1} E \left(\sum_i n_i \mathbf{u}_i \mathbf{u}_i^T \right) \\ \hat{\boldsymbol{\Sigma}}_\beta &= \left(\sum_i n_i \right)^{-1} E \left(\sum_{ij} \beta_{ij} \beta_{ij}^T \right) \\ \hat{s}^2 &= \left(\sum_i n_i \right)^{-1} E \left(\sum_i (\mathbf{r}_i - \mathbf{X}_{2i} \boldsymbol{\alpha}_2 - \mathbf{1}_{n_i} \mathbf{b}^T \mathbf{u}_i)^T (\mathbf{r}_i - \mathbf{X}_{2i} \boldsymbol{\alpha}_2 - \mathbf{1}_{n_i} \mathbf{b}^T \mathbf{u}_i) \right)\end{aligned}$$

The E and M steps are repeated until convergence is achieved. The convergence criteria is based on change in the expectation of the complete data log likelihood. Initial parameter estimates are found by fitting a two-stage model. Subject and pattern level random effects are estimated from a generalized mixed effects model for the ordinal outcome. The estimated pattern random effects are then fitted as fixed covariate in the dropout time model. Standard errors for all the parameters are obtained via parametric bootstrapping.

To test the proportional odds assumption of the proportional odds model, the z_{ijk} by I interaction is added in the I -th cumulative probability logistic regression. Corresponding indicator variables are added in the design matrix for β_{ij} . The coefficients of the interaction terms and the bootstrap standard errors are estimated. If all the interaction coefficients are close to zero, then the proportional odds model assumption is satisfied.

4. Application to the ESRD Study

We applied the random pattern mixture model to a sample 476 subjects from the ESRD study [23]. The study consists of the ESRD patients in the Fresenius Medical Corporation clinical dataset undergoing hemodialysis. The first 6 months were used as the baseline period in order to calculate their iron exposure, followed by observations every 3 months till 27 months or death. The ordinal anemia outcome was defined based on the blood hemoglobin. Level 1 is not anemic with hemoglobin greater than or equal to 13 for men and greater than or equal to 12 for non-pregnant women; level 2 is mild anemia with hemoglobin level between 9.3 and 13; and level 3 is moderate and severe anemia with hemoglobin level less than 9.5. Baseline iron treatment was defined as receiving 50mg or more of iron supplementation during the 6 month baseline period. Approximately two-thirds ($n = 298$) patients received iron at baseline. A total of 164 (34.5%) patients died during the follow up period with an average survival time of 14.9 months; 99 (33.2%) patients who received baseline iron died as compared to 65 (36.5%) patients who did not receive iron. Years of

ESRD, and some demographic characteristics such as age, gender, and race were also measured at baseline. Albumin levels were measured at each follow up visit. The objective of the analysis was to examine the effect of the 6 month cumulative iron exposure on anemia severity, accounting for the deaths in possibly the sickest patients.

Age was used as a latent variable to stratify the subjects into 10 patterns based on deciles. Age was used as the pattern variable since it led to approximate conditional independence of the longitudinal severity of anemia outcome and the time to dropout. It was also a clinical candidate as a surrogate for underlying disease severity, with older subjects tending to be sicker and being more likely to die. The percent of dropout, average survival time and average probability of anemia by pattern are shown in Table 1. The pattern with older subjects tends to have shorter survival time, lower probability of no anemia, and higher probability of being moderately or severely anemic. The marginal spearman correlation coefficients are 0.83 and -0.85 for the time to dropout and the average probabilities of no anemia and moderate or severe anemia separately. The partial spearman correlation coefficients are 0.20 and -0.30 , suggesting that the time to dropout and the ordinal anemia outcome are less related after stratifying the data by age. Thus, within each age pattern, severity of anemia and dropout time are independent, and the information between subjects within the same pattern can be borrowed from each other.

Three models were applied to estimate the effect of baseline iron on the severity of anemia: a proportional odds model with no link between anemia and the time to dropout, a shared parameter model linking the two models at the subject level, and a random pattern mixture model using age group as the random pattern variable. The covariates included in the longitudinal proportional odds models for anemia severity were baseline iron, time and albumin. The subject level random slope was essentially 0 and therefore only a subject level random intercept was included in all models. Time to dropout was log transformed and fitted with a linear mixed model with covariates of baseline iron and length of ESRD. The results for the three models are shown in Table 2.

The estimated cumulative odds ratio of having anemia $I(I=1,2)$ in the baseline iron exposure group vs. no iron was 1.26 ($p=0.15$) in the proportional odds model, 1.23 ($p=0.06$) in the shared parameter model and 1.31 ($p=0.01$) in the random pattern mixture model. The estimated cumulative odds ratios for each unit increase of albumin were 5.37, 4.81 and 4.57 in the three models respectively. The three models gave the same estimate for the time effect on anemia severity, with a cumulative odds ratio of 1.04 for each month increase. Similar estimates of iron, age and gender effects on the time to dropout were also obtained in the three models. Both joint models provided similar variance estimates for the subject level random effect. Compared to the proportional odds model with no link to the dropout process, the shared parameter model showed a marginally significant iron effect, while the random pattern mixture model estimated a larger and significant iron effect. The small link parameter of 0.10 ($p=0.001$) in the shared parameter model suggests that the dropout is almost noninformative at the individual subject level and the subject itself may not be a good surrogate for the dropout process. The greater link parameter of 0.89 ($p<0.001$) in the random pattern mixture model suggests that the dropout process may become informative at the pattern level. The random pattern mixture model can therefore correct some bias by

linking anemia severity to the dropout time at age pattern level. A sample of the fitted and observed values from the random pattern mixture model are shown in Figure 1. The robustness of the random pattern mixture model was tested in simulations.

5. Simulations

Simulations were performed to study the sensitivity of various model assumptions. The random pattern mixture model and the shared parameter model were compared by fitting simulated data from different true underlying models and censoring rates. In each simulation, 200 subjects were simulated with 500 data replicates generated. The study exit time was calculated so that a uniform percentage of subjects had a dropout time greater than the study exit time; giving a consistent censoring rate from one data replicate to the next. If a subject had dropout time greater than the study exit time, then this subject was censored and their last time was the study exit time; otherwise, the subject dropped out at the simulated dropout time. The longitudinal repeated outcomes were then simulated based on the dropout time or censor time. Longitudinal measurements were simulated with equi-spaced time intervals. Thus, the number of repeated observations for each subject ranged from two to twelve.

A proportional odds model for a 3 level longitudinal ordinal outcome was simulated. The first two simulations were performed assuming the data came from the true random pattern mixture model below, with 30% and 50% censoring rates respectively.

$$\begin{aligned} \text{logit}(\mu_{ijkl}) &= -3.2 + 2.3(I_{l=2}) + 2(\text{iron}) + 0.2(\text{time}) + \beta_{ij1} + (\text{time})\beta_{ij2} + u_i, \quad l=1, 2 \\ \log(r_{ij}) &= 0.6 + 0.3(\text{iron}) + 0.2(\text{length of ESRD}) + 1.4u_i + \varepsilon_{ij} \end{aligned}$$

where

$$\begin{aligned} \begin{pmatrix} \beta_{ij1} \\ \beta_{ij2} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}\right), \\ u_i &\sim N(0, 1.2), \\ \varepsilon_{ij} &\sim N(0, 0.7), \end{aligned}$$

and μ_{ijkl} is the cumulative probability of having the outcome l at the k -th time point for the j -th subject within the i -th pattern. A random pattern mixture model using the true pattern variable and a shared parameter model were fitted. Only the results from the main parameter of interest, longitudinal iron, are shown in Table 3. As expected, the random pattern mixture model using the true pattern variable provided unbiased estimates, while the shared parameter model had biased estimates due to the model misspecification. The shared parameter model gives substantial 8% bias in the longitudinal iron effect, and MSE is about 50% higher compared to the random pattern mixture model. Additionally, while the variance of the subject random effects are over estimated in an attempt to incorporate the missing pattern random effect, the shared parameter model cannot correct all the bias when the outcome and the informative dropout are actually linked at the pattern level. Thus, like

traditional longitudinal models, fitting a shared parameter model that ignores true clustering effects will potentially result in biased estimates.

The second two simulations were performed assuming the true underlying model was the shared parameter model below with 30% and 50% censoring rates respectively.

$$\begin{aligned} \text{logit}(\mu_{jkl}) &= -3.4 + 2.3(I_{t=2}) + 2(\text{iron}) + 0.1(\text{time}) + \beta_{j1} + (\text{time})\beta_{j2}, \quad l=1, 2 \\ \log(r_j) &= 1.2 + 0.13(\text{iron}) + 0.1(\text{length of ESRD}) + \beta_{j1} + 0.1\beta_{j2} + \varepsilon_j \end{aligned}$$

where

$$\begin{aligned} \begin{pmatrix} \beta_{j1} \\ \beta_{j2} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}\right), \\ \varepsilon_j &\sim N(0, 0.7), \end{aligned}$$

and μ_{jkl} is the cumulative probability of having the outcome l at the k -th time point for the j -th subject. Two random pattern mixture models with different pattern assignments were fitted as well as a shared parameter model. The first pattern assignment was based on the time to dropout. Subjects with the same dropout time were grouped into one pattern. In the second random pattern mixture model, the subjects were grouped into 10 equal sized patterns based on the length of ESRD. The results from the shared parameter model and two random pattern mixture models were compared, as shown in Figure 2 and Figure 3. The summary for the longitudinal iron parameter is shown in Table 3. Not surprisingly, the shared parameter model has unbiased results. The random pattern mixture model using the dropout time pattern had biased results for all the parameters. The longitudinal iron effect has more than 5% bias with 30% censor rate, and about 10 times of MSE with 50% censor rate. However the random pattern mixture model using length of ESRD pattern corrected the bias for the iron effects on both responses. It has only 1% bias for the longitudinal iron effect. It shows that even when the true underlying model is a shared parameter model, fitting a random pattern mixture model using an appropriate pattern variable can still have decent estimates. Time to dropout is not a good surrogate for the dropout process and a pattern variable based on a baseline covariate is a better choice when the true pattern is actually individual subjects.

One final simulation was performed where the true underlying model had a single subject level shared random intercept with a 50% censor rate. Both the shared parameter model with random intercept and slope and the random pattern mixture model with patterns based on ESRD were fitted to the simulated data. This gives a similar setting to simulation 4, that had the least biased estimates for both models, but in this case, both fitted models were misspecified. Table 3 summarizes the properties of the longitudinal iron parameter in this simulation. The random pattern mixture model has smaller bias and gains efficiency compared to the shared parameter model. When both models are misspecified, the random pattern mixture model is still less biased and more efficient at estimating the parameter of interest. Similar results were obtained for the other fixed effects parameters in the models. In the application context, the random pattern mixture model is powerful and more likely to

detect the iron effect compared to the shared parameter model. The shared parameter model has bias towards null and bigger MSE if the model is misspecified. The iron effect tends to be not significant. If the random pattern model or even both models are misspecified, the random pattern model has smaller bias and are more efficient at estimating the iron effect with an appropriate pattern variable.

6. Conclusions

We have introduced the random pattern mixture model for longitudinal ordinal outcomes with informative dropouts. The random pattern mixture model assumes that conditional on the pattern variable the longitudinal outcome is independent of the dropout process. The patterns are based on stratification of a surrogate variable of the dropout process. In the absence of a good surrogate, time-to-dropout can be used to create patterns. However, in the situation with large percentage of dropouts, all the dropouts are grouped as one single pattern, which can reduce efficiency. The study model allows subjects within the same pattern to borrow information from each other and is considered more flexible than a shared parameter model when there are multiple subjects per pattern. When there are very few subjects per pattern, the random pattern mixture model approaches the shared parameter model and the improved efficiency may be limited. We applied the model to a dataset from the ESRD study, analyzing the effect of baseline iron on the longitudinal, ordinal, severity of anemia outcome. Age was found to be a good surrogate for the dropout process and was used as a pattern variable to stratify the subjects into 10 patterns. The model estimated an increased and significant iron effect compared to a proportional odds model and a shared parameter model. Simulations showed that the random pattern mixture model is robust under various assumptions and model misspecification. The random pattern mixture model can fit the data well with a carefully chosen pattern variable even when the true underlying model is a shared parameter model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract/grant sponsor: National Institutes of Health Grant R01 GM104470

References

1. McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B.* 1980; 42:109–142.
2. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics.* 1994; 50:933–944. [PubMed: 7787006]
3. Guo W, Ratcliffe SJ, TenHave TR. A random pattern-mixture model for longitudinal data with dropouts. *Journal of the American Statistical Association.* 2004; 99(4):927–937.
4. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics.* 1988; 44:175–188.
5. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics.* 1989; 45:939–955. [PubMed: 2486189]

6. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*. 1992; 11:1861–1870. [PubMed: 1480878]
7. DeGruttola V, Tu XM. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*. 1994; 50:1003–1014. [PubMed: 7786983]
8. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*. 1993; 88:125–134.
9. Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika*. 1994; 81:471–483.
10. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics*. 1995; 51:151–168. [PubMed: 7766771]
11. TenHave TR, Kunselman AR, Pulkstenis EP, Landis JR. Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics*. 1998; 54:367–383. [PubMed: 9544529]
12. Pulkstenis EP, TenHave TR, Landis JR. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*. 1998; 93:438–450.
13. Albert PS, Follmann DA. Modeling repeated count data subject to informative dropout. *Biometrics*. 2000; 56:667–677. [PubMed: 10985201]
14. Birmingham J, Fitzmaurice GM. A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics*. 2002; 58:989–996. [PubMed: 12495154]
15. TenHave TR, Reboussin BA, Miller ME, Kunselman A. Mixed effects logistic regression models for multiple longitudinal binary functional limitation responses with informative drop-out and confounding by baseline outcomes. *Biometrics*. 2002; 58:137–144. [PubMed: 11890309]
16. Gallop RJ, TenHave TR, Crits-Christoph P. A mixed effects markov model for repeated binary outcomes with non-ignorable dropout. *Statistics in Medicine*. 2006; 25:2398–2426. [PubMed: 16252275]
17. TenHave TR, Miller ME, Reboussin BA, James MK. Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics*. 2000; 56:279–287. [PubMed: 10783807]
18. Pulkstenis E, TenHave TR, Landis JR. A mixed effects model for the analysis of ordinal longitudinal pain data subject to informative drop-out. *Statistics in Medicine*. 2001; 20:601–622. [PubMed: 11223903]
19. Chen LM, Ibrahim JG, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine*. 2011; 30(18):2295–2309. doi:10.1002/sim.4263. [PubMed: 21590793]
20. Feldman HI, Santanna J, Guo W, Furst H, Franklin E, Joffe M, Marcus S, Faich G. Iron administration and clinical outcomes in hemodialysis patients. *Journal of the American Society of Nephrology*. 2002; 13:734–744. [PubMed: 11856779]
21. Dempster A, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39:1–38.
22. Pinheiro J, Chao E. Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear models. *Journal of Computational and Graphical Statistics*. 2006; 15(1):58–81.
23. Ratcliffe SJ, Guo W, TenHave TR. Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*. 2004; 60:892–899. [PubMed: 15606409]

Figure 1.
Profile plots of random subjects from the ESRD Study. Solid line with + = observed anemia category; dashed line with O = predicted anemia category from the random pattern mixture model.

Figure 2.

Simulation results assuming that the true model is a shared parameter model, censor rate=30%. 1, shared parameter model. 2, random pattern mixture model using dropout time pattern. 3, random pattern mixture model using length of ESRD pattern.

Figure 3.

Simulation results assuming that the true model is a shared parameter model, censor rate=50%. 1, shared parameter model. 2, random pattern mixture model using dropout time pattern. 3, random pattern mixture model using length of ESRD pattern.

Table 1

Description Statistics by Age Pattern.

Pattern	n	Death	Average survival time (months)	Average probability of no anemia	Average probability of moderate or severe anemia
1	58	10.3	21.53	18.57	17.35
2	41	14.6	24.88	17.43	21.14
3	67	23.9	23.33	15.43	25.46
4	40	25.0	24.23	15.72	25.47
5	51	35.3	20.56	6.65	41.55
6	51	39.2	20.87	8.02	40.64
7	49	42.9	20.11	8.90	39.88
8	45	48.9	20.62	5.14	38.26
9	37	62.2	20.05	4.28	49.03
10	37	59.5	18.81	4.69	51.17

Table 2

Results from modeling the ESRD data, assuming proportional odds model with no link between outcome and time to dropout, shared parameter model, and random pattern mixture model based on age. Parameter estimates and bootstrapped standard errors are shown.

Parameter	Proportional odds model	Shared parameter model	Random pattern mixture model
Model for anemia outcome			
Intercept1	−10.05 (0.53)	−9.43 (0.51)	−9.41 (0.53)
Intercept2	4.03 (0.10)	3.82 (0.09)	4.03 (0.09)
Iron	0.23 (0.16)	0.21 (0.11)	0.27 (0.11)
Time	0.04 (0.005)	0.04 (0.004)	0.04 (0.004)
Albumin	1.68 (0.13)	1.57 (0.12)	1.52 (0.13)
σ_{β}^2	2.03 (0.20)	1.52 (0.11)	1.86 (0.14)
σ_u^2			0.17 (0.05)
Model for time to dropout			
Intercept	3.49 (0.11)	3.54 (0.08)	3.61 (0.08)
Iron	0.10 (0.10)	0.11 (0.08)	0.13 (0.09)
Length ESRD	0.02 (0.01)	0.02 (0.01)	0.002 (0.01)
b		0.10 (0.03)	0.89 (0.11)
σ_s^2		0.91 (0.04)	0.74 (0.06)

Table 3

The average properties of the longitudinal iron parameter estimates across the 5 simulation settings.

Simulation	Censor Rate	Linked at: True Link	Shared Parameter Model Subject Level			Random Pattern Mixture Model ESRD Pattern			Random Pattern Mixture Model Dropout Time Pattern		
			% Bias	MSE	RE	% Bias	MSE	RE	% Bias	MSE	RE
1	30%	ESRD Pattern	-7.99	0.092	0.66	0.57	0.061	1			
2	50%	ESRD Pattern	-7.25	0.076	0.67	0.29	0.051	1			
3	30%	Subject Level	1.57	0.050	1	1.23	0.066	0.76	5.30	0.085	0.59
4	50%	Subject Level	-0.19	0.057	1	-1.31	0.050	1.14	-0.12	0.51	1.12
5	50%	Subject Intercept	-6.84	0.103	0.53	-1.66	0.083	1			

MSE=Mean square error; RE = Relative efficiency.