Published in final edited form as:

Stat Med. 2014 September 10; 33(20): 3434–3452. doi:10.1002/sim.5596.

Bayesian multivariate growth curve latent class models for mixed outcomes

Benjamin E. Leiby^{1,*}, Thomas R. Ten Have², Kevin G. Lynch³, and Mary D. Sammel²

¹Division of Biostatistics, Thomas Jefferson University, 1015 Chestnut St. Suite M100, Philadelphia, PA, 19107

²Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 6th Floor Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104

³Department of Psychiatry, University of Pennsylvania School of Medicine

SUMMARY

In many clinical studies, the disease of interest is multi-faceted, and multiple outcomes are needed to adequately capture information about the characteristics of the disease or its severity. In analysis of such diseases, it is often difficult to determine what constitutes improvement due to the multivariate nature of the outcome. Furthermore, when the disease of interest has an unknown etiology and/or is primarily a symptom-defined syndrome, there is potential for the disease population to have distinct subgroups. Identification of population subgroups is of interest as it may assist clinicians in providing appropriate treatment or in developing accurate prognoses. We propose multivariate growth curve latent class models that group subjects based on multiple symptoms measured repeatedly over time. These groups or latent classes are defined by distinctive longitudinal profiles of a latent variable which is used to summarize the multivariate outcomes at each point in time. The mean growth curve for the latent variable in each class defines the features of the class. We develop this model for any combination of continuous, binary, ordinal or count outcomes within a Bayesian hierarchical framework. Simulation studies are used to validate the estimation procedures. We apply our model to data from a randomized clinical trial evaluating the efficacy of Bacillus Calmette-Guerin in treating symptoms of Interstitial Cystitis where we are able to identify a class of subjects for whom treatment is effective.

1. Introduction

In clinical studies the outcomes of interest are often measured using different metrics. For example, in studies of Interstitial Cystitis (IC), multiple measures are recorded to describe disease severity. Two of these symptoms, urinary pain and urgency to void, are measured on a ten-point (0–9) Likert scale and are often analyzed as normally-distributed variables or are collapsed into ordinal severity categories. A third symptom, frequency of voiding, is typically modeled as a poisson-distributed variable. Symptoms scales and other global assessments of symptoms have also been used to describe disease status, but none gives a complete picture. Consequently, joint analysis of these multiple outcomes measured using

*Correspondence to: Division of Biostatistics, Thomas Jefferson University, 1015 Chestnut St. Suite M100, Philadelphia, PA 19107.

different metrics is desirable for a better understanding of the full spectrum of the underlying disease process. Patients diagnosed with IC are also heterogeneous in their symptom presentation. While some patients will experience all three symptoms, others may only have elevated values in only one or two domains. The different patterns of symptoms may indicate different disease processes at work and potentially different responses to treatment. The ability to group subjects based on their pattern of symptom expression may aid in the understanding and effective treatment of complex diseases.

In previous work [1], we incorporated the longitudinal factor analytic model of Roy and Lin [2] into a latent class structure to jointly modeled the association among two or more continuous, normally-distributed variables while accounting for heterogeneity in the symptom trajectories over time. We applied this model to the two most closely-related symptom measurements, pain and urgency, and found a group of subjects who exhibited improvement over time and whose improvement was greater when treated with *Bacillus Calmette-Guerin* (BCG). In this paper we propose an extension of our previous model to include outcomes which are measured using different metrics, such as categorical or count outcomes.

Two primary approaches have been proposed for factor analytic models with categorical manifest variables. The first approach, for the case of binomial or ordinal observed variables, imposes the assumption of an underlying normally-distributed variable for each observed variable. The probability that the observed variable is equal to a given value is defined by the probability that the underlying variable is between two thresholds. The correlation structure of these underlying variables are then modeled using the usual factor analytic model. This is the probit model approach which was developed by Muthén([3, 4, 5]) and is incorporated in Mplus software. Arminger and Kusters [6] and Dunson [7] used this approach in a Bayesian model. While attractive due to relative ease of estimation, the underlying variable approach is limited to observed variables that can be easily linked to underlying continuous variables and excludes common types of categorical outcomes such as counts.

The second approach is the so-called response function approach where the observed categorical variable is linked directly to the latent variable through some functional relationship without the use of an intermediary, underlying continuous variable for each observed variable. This approach is most fully developed for the cross-sectional setting by Sammel et al [8] and Moustaki and Knott [9] where the manifest variables can be any member of a one-parameter exponential family. By utilizing the conditional independence assumption conditional on the latent variables, this framework readily accommodates variables of mixed distribution types and is not limited to binary and ordinal manifest variables.

Dunson [10] developed a latent variable model for repeated measurements of mixed outcomes (binary, count, ordinal, and continuous). This model is similar to the model of Roy and Lin in that the focus is on evaluating the associations between covariates of interest and the latent variable, with the correlation among the outcomes over time being considered as a nuisance. Both Roy and Lin and Dunson assume that the underlying trait(s) of interest, the

latent variable(s), is (are) continuous and normally distributed and that there is a homogeneous latent trait; that is, subjects belong to a single class. An alternative approach is to assume a discrete latent variable (latent classes). For example, Miglioretti [11] developed a latent transition model where multiple outcomes are assumed to be manifestations of an underlying unobserved categorical disease variable where the levels of the variable correspond to unobserved disease “states”. Correlation among the repeated measurements of the observed variables over time is modeled by allowing subjects to transition from one state to another with certain transition probabilities. Such a model may not appropriately capture gradual changes in disease status over short observation periods.

In this paper, we propose a multivariate growth curve latent class model that groups subjects based on multiple symptoms measured repeatedly over time. These groups or latent classes are characterized by distinctive longitudinal profiles of a latent variable which is used to summarize the multivariate outcomes at each point in time. The mean growth curve for the latent variable in each class is the primary defining feature of the class. We develop this model for any combination of continuous, binary, ordinal or count outcomes from an exponential family within a Bayesian hierarchical framework. We specify a Bayesian hierarchical model and develop a Markov Chain Monte Carlo (MCMC) algorithm to estimate the posterior distributions of the model parameters. Our model differs from the work of Roy and Lin in that it accommodates outcomes of mixed types and allows for multiple latent classes of trajectories. It differs from the work of Dunson in assuming a single underlying continuous trait, in allowing for multiple classes, and in its approach to modeling the correlation among repeated measures over time. Finally, in contrast to the work of Miglioretti, our model assumes subjects belong to a single class at baseline with no transition between classes over time.

This paper is organized as follows: Sections 2 through 4 present the proposed model and describe the method of estimation. In section 5 we present results from application of this proposed model to data from a randomized trial in IC. Section 6 presents results of a simulation study to validate our estimation procedure. Finally, section 7 discusses our results and suggestions for future work.

2. Model Specification

Begin by defining Y_{ijm} be the m^{th} observed response variable ($m = 1, \dots, M$) measured at the j^{th} time point ($j = 1, \dots, n_i$) for the i^{th} subject ($i = 1, \dots, N$). Let C_i to be the latent class membership of subject i with $C_i = k$ if subject i belongs to class k ($k = 1, \dots, K$) and let $(c_{i1}, \dots, c_{iK})$ be a corresponding vector of indicator variables where $c_{ik} = 1$ if $C_i = k$ and $c_{ik} = 0$ otherwise. Then, conditional on $C_i = k$, Y_{ijm} follows a generalized linear factor analytic model such that

$$g(E(Y_{ijm}|C_i=k)) = \alpha_{mk} + \lambda_{mk}Z_{ij} + b_{im} \quad (1)$$

with $b_{im} \sim N(0, \tau_m^2)$ to describe the correlation over time for the m^{th} outcome. We use the canonical link functions so that $g(\cdot)$ is the identity link for normal outcomes, the log link for Poisson outcomes, and the logit link for binary outcomes. For normal outcomes, we assume

the residual variance terms $Y_{ijm} - E(Y_{ijm}|C_i=k) = e_{ijm}^{(y)} \sim N(0, \sigma_m^2)$. Conditional on the latent variable Z_{ij} , the M outcomes at each time point are assumed to be independent, and conditional on Z_{ij} and b_{im} , the n_i repeated measures of the m^{th} outcome are also independent. The parameters (α_{mk}) and (λ_{mk}) are the outcome-specific intercept and factor loading which can vary by class but are static with respect to time. The factor loading parameters express the strength of association between the latent variable and the manifest variables and are a function of the mean and correlation among the manifest variables. Allowing the loadings to vary by class admits the possibility of differing types of patients whose disease manifests with different symptoms. The assumption of static factor loadings across time is somewhat restrictive in that it assumes that the way the latent disease state manifests itself through the various symptoms remains the same over time. This may not be reasonable in some settings (e.g., situations of multiple year follow-up), but is likely to be acceptable in the situation of a relatively short-term clinical trial.

While the logit link is preferable to the probit link as it provides a natural link to the odds ratio, difficulties in estimation exist when the latent variable and random effects are normally distributed. For binary outcomes, an equivalent representation of the logit model can be made through the assumption of an underlying latent variable which follows a scale mixture of normal distributions [12]. Let Y_{ijm}^* be an unobserved continuous variable with $Y_{ijm} = 2$ if $Y_{ijm}^* > 0$ and $Y_{ijm} = 1$ if $Y_{ijm}^* \leq 0$. Further,

$$Y_{ijm}^* = \alpha_{mk} + \lambda_{mk} Z_{ij} + b_{im} + e_{ijm}^{y*} \quad (2)$$

$$e_{ijm}^{y*} \sim N(0, \sigma_{ijm}^2) \quad (3)$$

$$\sigma_{ijm}^2 = 4 * \psi_{ijm}^2 \quad (4)$$

$$\psi_{ijm} \sim KS \quad (5)$$

where KS is the Kolmogorov-Smirnov distribution. Integrating out ψ_{ijm} yields a logistic distribution for e_{ijm}^{y*} and a mixed effects logistic regression model for the observed outcome Y_{ijm} . Specification of the model in this way allows us to retain conjugacy of prior and random effects distributions which simplifies estimation of the posterior distribution [13].

To accommodate ordinal outcomes, we extend this framework by specifying a set of threshold parameters $\alpha^{(0)}, \dots, \alpha^{(D)}$ so that for $d = 1, \dots, D_m$, $Y_{ijm} = d$ if $\alpha_{mk}^{(d-1)} < Y_{ijm}^* < \alpha_{mk}^{(d)}$ where $-\infty = \alpha_{mk}^{(0)} < \alpha_{mk}^{(1)} < \dots < \alpha_{mk}^{(D-1)} = 0 < \alpha_{mk}^{(D)} = \infty$ and Y_{ijm}^* is assumed to follow the same model as described in (2). Since λ_{mk} in (2) is constant across the levels of the ordinal variable, this parametrization leads to a mixed effects proportional odds logistic regression model conditional on the latent variables Z_{ij} and C_i .

Next, conditional on class, $C_i = k$, Z_{ij} follows a mixed effects model:

$$Z_{ij} = \mathbf{X}_{ij}^{(Z)T} \boldsymbol{\eta}_i + e_{ij}^{(Z)} \quad (6)$$

$$\boldsymbol{\eta}_i = \boldsymbol{\gamma}_k + a_i \quad (7)$$

where $e_{ij}^{(Z)} \sim N(0, 1)$, $\mathbf{X}_{ij}^{(Z)}$ is an $r \times 1$ vector of observed covariates for Z_{ij} and $\boldsymbol{\gamma}_k$ is the mean of $\boldsymbol{\eta}_i$, the vector of coefficients. Since we are modeling the trajectory of the latent variable over time, Z_{ij} includes function(s) of time j . In addition, assuming $a_i \sim N(0, \Psi_a)$ is used to specify the random effects while fixed effects are specified by setting the appropriate variance components equal to 0.

Finally, the probability of class membership, π_{ik} , is modeled using a reference category multinomial logistic regression model:

$$Pr(C_i = k) = \pi_{ik} = \frac{e^{\mathbf{X}_i^{(C)T} \boldsymbol{\delta}_k}}{\sum_{k=1}^K e^{\mathbf{X}_i^{(C)T} \boldsymbol{\delta}_k}} \quad (8)$$

where $\boldsymbol{\delta}_K \equiv \mathbf{0}$ and $\mathbf{X}_i^{(C)}$ is a $p \times 1$ vector of observed subject-level covariates. For model identifiability, λ_{1k} is constrained to be positive, and the mean of the intercept in the model for Z_{ij} is fixed at 0.

3. Prior Distributions

We specify normal priors for the outcome-specific intercepts, factor loadings, and fixed effects parameters in the mixed effects model. These priors may be defined separately for each class, or the same priors may be used for all classes. In practice, we use the data to inform our choice of centering values for the priors of the intercepts, factor analysis and regression models to center the priors for the factor loadings and fixed effects. Specifically,

$$\alpha_{mk} \sim \text{Normal}(\alpha_m^*, \xi_{\alpha_m}) \quad (9)$$

$$\lambda_{mk} \sim \text{Normal}(\lambda_m^*, \xi_{\lambda_m}) \quad (10)$$

$$\boldsymbol{\gamma}_k \sim \text{Normal}(\boldsymbol{\gamma}^*, \Phi) \quad (11)$$

where, for a one-class model, α_m^* are the intercepts from fitting linear or nonlinear mixed effects models to the entire data set. We set $\lambda_m^* = 1$ implying the same association between the latent variable and each observed variable of the same type, and we set $\boldsymbol{\gamma}^* = \mathbf{0}$ to indicate prior belief of no trend in the latent variable over time. For models with more than one class, α_m^* , λ_m^* , and $\boldsymbol{\gamma}^*$ are the posterior medians from the one-class model. To insure that the priors are noninformative, we specify large variances. For the one-class model we assume $\xi_{\alpha_m} = 100$, $\xi_{\lambda_m} = 100$, and $\Phi = 100 * I(\dim(\boldsymbol{\gamma}))$. For models with more than one class, prior variances are specified as N times the variance of the posterior distribution from the one-

class model. For the threshold parameters of the ordinal outcomes, we specify an ordered normal prior distribution where each $\alpha_{mk}^{(d)} \sim N(\alpha_{mk}^{(d)*}, 100)$ for $d = 1, \dots, D_m - 2$ subject to $\alpha_{mk}^{(1)} < \alpha_{mk}^{(2)} < \dots < \alpha_{mk}^{(D_m-1)} = 0$. As for the intercept terms, the $\alpha^{(d)*}$ are the threshold parameters from fitting logistic mixed effects models to the entire data set.

We specify inverse-gamma priors for the variance of the outcome-specific random intercept in the factor analytic model (1) and for the residual variances of the continuous outcomes, e_{ijm}^y , and an inverse-Wishart prior for the covariance matrix of the random effects in (7).

$$\alpha_m^2 \sim \text{Inv-gamma}(r_{1m}, r_{2m}) \quad (12)$$

$$\tau_m^2 \sim \text{Inv-gamma}(r_{3m}, r_{4m}) \quad (13)$$

$$p(\Psi_a) \propto |\Psi_a|^{-(\dim(\Psi_a) + \nu + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(S\Psi_a^{-1})\right) \quad (14)$$

where $\nu = 1$, $S = \text{diag}(1)$ and $p(\cdot)$ indicates the probability density function. Setting $r_{1m} = r_{2m} = 0.5$ and $r_{3m} = r_{4m} = 0.5$ yields the inverse chi-square prior with 1 degree of freedom. If the amount of information available in the binary or ordinal outcomes is limited (i.e., low prevalence of one category at one or more time points), it may be necessary to specify more informative priors which can be achieved by specifying larger values of r_{3m} and r_{4m} . Alternatively, setting $r_{1m} = -0.5$ and $r_{2m} = 0$ results in the improper noninformative uniform prior on σ recommended by Gelman [14].

For class probability models with covariates, we specify $\delta_k \sim \text{Normal}(0, \frac{9}{4}I)$ [15]. When we specify an intercept-only model for the class probability, we assumed a Dirichlet prior with equal prior sample sizes $\omega_1^*, \dots, \omega_K^* = \omega^*$ for π_k .

4. Estimation of the Posterior Distribution

Without loss of generality, let the first m_1 manifest outcomes be normally distributed, the next m_2 outcomes be ordered categorical, the next m_3 outcomes be binary, and the remaining outcomes follow a Poisson distribution, and consider $\beta_{mk} = (\alpha_{mk}, \lambda_{mk})^T$. The full joint posterior distribution is given by

$$p(\Theta|y) \propto p(y|\Theta)p(\Theta) \quad (15)$$

where

$\Theta = [\beta_1, \dots, \beta_K, \alpha_1^{(\bullet)}, \dots, \alpha_K^{(\bullet)}, \gamma_1, \dots, \gamma_K, \delta, \sigma_1^2, \dots, \sigma_{m_1}^2, \tau_1^2, \dots, \tau_M^2, \Psi_a, \mathbf{y}^*, \mathbf{b}, \mathbf{z}, \mathbf{a}, \mathbf{C}]$ and includes the parameters specified in section 2 as well as the latent variables, latent class assignments, and the random effects.

Through assumptions of conditional independence, the likelihood can be written as a product over N subjects, K classes, n_i observations per subject, and M outcomes as

$$p(\mathbf{y}|\Theta) = \prod_{i=1}^N \left\{ \prod_{k=1}^K \left[\prod_{j=1}^{n_i} \left[\prod_{m=1}^{m_1} p(y_{ijm}|\beta_{mk}, z_{ij}, b_{im}, C_i=k, \sigma_m^2) \right. \right. \right. \\ \times \prod_{m=m_1+1}^{m_1+m_2} p(y_{ijm}|\alpha_{mk}^{(\cdot)}, y_{ijm}^*, C_i=k) \\ \times \prod_{m=m_1+m_2+1}^{m_1+m_2+m_3} p(y_{ijm}|y_{ijm}^*, C_i=k) \\ \left. \left. \left. \times \prod_{m=m_1+m_2+m_3+1}^M p(y_{ijm}|\beta_{mk}, z_{ij}, b_{im}, C_i=k) \right] \right] \right]^{c_{ik}} \right\}$$

Similarly, the prior distribution of the random effects and the parameters can be written as:

$$p(\Theta) = \prod_{i=1}^N \left\{ \prod_{k=1}^K \left[\prod_{j=1}^{n_i} \left[\prod_{m=m_1+1}^{m_1+m_2} p(y_{ijm}^*|\beta_{mk}, z_{ij}, b_{im}, C_i=k) \prod_{m=m_1+m_2+1}^{m_1+m_2+m_3} p(y_{ijm}^*|\beta_{mk}, z_{ij}, b_{im}, C_i=k) \right. \right. \right. \\ \times p(z_{ij}|\gamma_k, \mathbf{x}_{ij}^Z, \mathbf{a}_i, C_i=k) \left. \left. \left. p(C_i=k|\mathbf{x}^C, \delta) \right] \right]^{c_{ik}} \prod_{m=1}^M p(b_{im}|\tau_m^2) p(\mathbf{a}_i|\Psi_a) \right\} \\ \times \prod_{k=1}^K \left[p(\gamma_k) \prod_{m=1}^M p(\alpha_{mk}) p(\lambda_{mk}) \prod_{m=m_1+1}^{m_1+m_2} p(\alpha_{mk}^{(\cdot)}) \right] \\ \times \prod_{k=1}^{K-1} p(\delta_k) \prod_{m=1}^{m_1} p(\sigma_m^2) \prod_{m=1}^M p(\tau_m^2) p(\Psi_a).$$

Simulation of the posterior distribution proceeds via a Gibbs sampler where each parameter is sampled from its full conditional distribution. When the full conditional distribution is not available in closed form draws proceed via Metropolis or rejection sampling. For continuous outcomes, all parameters may be drawn directly from their full conditional distribution. For count outcomes a Metropolis algorithm is needed to sample from the full conditional distribution of β_{mk} and b_{im} . For ordinal outcomes a Metropolis algorithm is used to sample from the full conditional distribution of the threshold parameters [16]. For binary and ordinal outcomes, the variance of y_{ijm}^* , σ_{ijm}^2 is drawn using rejection sampling [13]. When count outcomes are included, the latent variable (z_{ij}) is drawn via a Metropolis algorithm. When covariates are included in (8), a Metropolis algorithm is used to draw the logistic regression parameters for the class assignments. Full details of the Gibbs sampler may be found in the appendix.

Three MCMC chains for each model were run with subjects assigned at random to different classes at the beginning of each chain. Starting values for the intercepts, factor loadings, and fixed effects parameters in the mixed effects model for the latent variables were varied by chain. In a one-class model, starting values for the intercept terms, threshold parameters, and variance parameters were based on fitting separate linear or nonlinear mixed effects models

for each outcome. Fixed effects for the latent variable were started at 0 and factor loadings were started at 1. For models with more than one class, starting values were based on results of the one-class model. For all models, we ran chains of 40,000 iterations, keeping every 20th iteration, until convergence was achieved for all parameters as assessed by the Gelman-Rubin statistic [17].

5. Application to Data from an Interstitial Cystitis Randomized Clinical Trial

In this section we present an application of our model and estimation method to data from a multi-center clinical trial to evaluate the efficacy of BCG in treating IC symptoms. Details of the clinical trial and results from the primary analysis have been published elsewhere [18]. Two hundred sixty-five (265) subjects were randomized to treatment with intravesical BCG or intravesical placebo. Symptom and other variables were recorded at baseline and at 4 subsequent visits (approximately 8, 18, 26, and 34 weeks after randomization). In this analysis we focus on five symptom measurements. Pain and urgency, both measured on 10 point (0-9) Likert scales, were measured at baseline and at each subsequent visit. For this analysis, we utilize a clinically relevant coding of these variables which collapses each symptom into a 3-level ordinal variables with levels 1-” mild” [0, 3], 2-” moderate” (3, 6] and 3-” severe” (6, 9]. Twenty-four hour frequency, the total number of voids during a 24-hour period, was collected at baseline and weeks 18, 26, and 34 and is analyzed as a poisson-distributed variable. The Wisconsin IC Symptom Inventory score is a summary symptom score with values ranging from 0–42 and was treated as a normally-distributed variable [19]. Finally, the Global Response Assessment (GRA) records each subject’s overall evaluation of their symptoms relative to baseline on a 7-point scale of markedly worse, moderately worse, slightly worse, no change, slightly improved, moderately improved and markedly improved. For this analysis we used a dichotomized version of the GRA. To maintain consistency of coding so higher values indicated greater severity of symptoms, moderate or marked improvement was coded as GRA= 1 and no improvement was coded as GRA= 2. The GRA is recorded at each of the follow-up visits but, by definition, is missing at baseline. A fixed value, GRA= 2, was assumed for all subjects at baseline. The GRA at week 34 was the primary outcome of the trial. In the study, 16% of patients overall (21% in the treatment group; 12% in the placebo group) indicated positive response, yielding a non-significant result ($p=0.062$).

In this analysis, the five symptoms are assumed to be manifestations of a single latent symptom severity measure which is assumed to follow a quadratic curve in time. A random intercept is included to allow for subject to subject differences in baseline severity. Fixed effects for treatment, age, and log base 10 transformed duration of symptoms as well as treatment by time and treatment by time² interaction terms were included as factors influencing severity. All covariates were centered at their overall sample means, and time was centered at week 18 so that intercept terms and the main effect of treatment indicate effects at week 18. We include quadratic terms to allow for possible worsening in symptoms after treatment was ended. For all parameters we employed the noninformative priors as described in section (3). Since the goal of our modeling is the identification of latent classes characterized by differences in symptom trajectories rather than baseline characteristics, we took a two-step approach to covariate inclusion. First, we fit the models without covariates

to identify the number of classes. Based on class assignments from the final model at this stage, we screened potential covariates for possible association with class membership. Any marginally significant differences (defined as $p < 0.20$) would then be included in a final run of the model.

5.1. Model Choice

To choose the number of classes, we used the Deviance Information Criterion (DIC) [20] which was based on the complete data likelihood, that is, $p(\mathbf{y}, \mathbf{z}, \mathbf{b}, \mathbf{c}|\boldsymbol{\theta})$, referred to as DIC_4 in [21] and defined

$$DIC = -4E_{\boldsymbol{\theta}, \mathbf{z}, \mathbf{b}, \mathbf{c}}[\log p(\mathbf{y}, \mathbf{z}, \mathbf{b}, \mathbf{c}|\boldsymbol{\theta})|\mathbf{y}] \quad (16)$$

$$+2E_{\boldsymbol{\theta}, \mathbf{z}, \mathbf{b}, \mathbf{c}}[\log p(\mathbf{y}, \mathbf{z}, \mathbf{b}, \mathbf{c}|E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{z}, \mathbf{c}])|\mathbf{y}] \quad (17)$$

where $\boldsymbol{\theta}$ is the vector of all parameters in the model. Celeux et al [21] determined that this version of the DIC was “most reliable” in picking the best model while avoiding the tendency to inadequately penalize complex models.

The first term (16) is easily approximated by evaluating the log complete likelihood at each iteration of the MCMC algorithm (i.e., draws of the parameters and the random effects and latent variables) and averaging across iterations. The second term (17) is also approximated by calculating the log complete likelihood at each iteration of the MCMC chain, however instead of using the values of the parameters for a given iteration, $E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{z}, \mathbf{c}]$ is needed at each iteration. To approximate this quantity, at each iteration of the MCMC algorithm, 200 additional iterations of the MCMC chain were run with $\mathbf{b}$, $\mathbf{z}$, and $\mathbf{c}$ fixed and the average of $\boldsymbol{\theta}$ over those 200 iterations was computed to approximate $E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{z}, \mathbf{c}]$. Using the complete data likelihood is preferred in this setting as it avoids the potential for DIC values smaller than $-2\log(\text{likelihood})$, a problem reported when using the observed data likelihood for mixture models [21].

To assess model fit we used posterior predictive checks ([11, 22]) by generating 1000 data sets from the posterior predictive distribution. We checked the fit of the continuous and count outcomes at each time point graphically through the use of QQ-plots. To assess the fit for the ordinal and binary outcomes, we calculated the number of subjects in each category by week, comparing the observed value to the posterior predictive distribution with a chi-square statistic

$$\chi_p^2 = \sum_{c=1}^C \frac{(O_{pc} - E_c)^2}{E_c} \quad (18)$$

where O_{pc} is the cell count for posterior predictive data set p and E_c is the cell count for the original data. When $E_c = 0$, O_{pc} was used in the denominator. This statistic was summarized by calculating the mean over the 1000 data sets by week and estimating

$Pr(\chi_p^2 > \chi_{0.95, df=C-1}^2)$. In order to determine how well our model was reproducing the

correlation structure of the data, we calculated the Spearman correlation coefficient for all pairs of observed variables.

5.2. Results

In this section, we report results for one- and two-class models. Table I gives the median and 95% credible intervals for the posterior distribution for all model parameters from the one- and two-class models. Figures 1 and 2 give the mean latent symptom curve for the placebo and treatment groups. When considering the one-class model results, we see that on average, subjects enrolled in the study improved with respect to the latent symptom score. This analysis indicates that subjects randomized to BCG had marginally better symptoms over time than those randomized to the placebo arm. The right hand side of the figure displays the difference in symptom severity between the two treatment arms. The credible interval crosses 0 at approximately 13 weeks indicating significant improvement for the treated group after that point in time.

Using a two-class model, we were able to identify 57% “non-responder” and 43% “responder” classes. The non-responder class, on the left, is characterized by slight improvement in the latent symptom score at the middle of the study with a return to baseline symptom levels by the end. The figure on the right depicts the responder class. Responders have a slightly higher latent symptom score at baseline, and the trend in symptoms is characterized by more substantial improvement in the latent symptom score by the middle of the study with only a slight worsening of symptoms after 20 weeks. In the non-responder class there was no difference in treatment arms with respect to the mean latent symptom curves. In the responder class, the BCG group had less severe symptoms throughout the study with statistically significant differences existing from before week 10 through the end of the study (see figure 3). Using the DIC criterion for selecting models, the two-class model is preferred over the one-class model ($DIC_2 = 23886$ vs. $DIC_1 = 23919$). A limited set of variables including age, race, sex, duration of disease, and treatment arm were screened as potential predictors of class membership. No covariates were identified as possible predictors of class membership, and so the two-class model with an intercept only class probability model was selected as our final model.

When considering the symptom-specific results for the two-class model (Figure 4), we see that model-estimated means at baseline were similar for non-responders and responders for the Wisconsin Total Score, pain and urgency. At week 8, the first time GRA was measured, placebo responders have similar GRA to non-responders while treatment responders are showing some improvement. In total, these results suggest that the treatment may be most effective at reducing pain and urgency with a more modest effect on frequency. In addition, patients with fewer daily voids at baseline may be more likely to respond.

Figure 5 shows representative QQ-plots for the Wisconsin Total Score and 24-hour frequency at week 0. Our model, with its assumption of normality, appears to provide a good fit to the data for the Wisconsin score. The model fits reasonably well for 24-hour frequency, although there is some under-estimation at both ends of the distribution. This may indicate that the poisson model with the random intercept may not be sufficiently overdispersed to adequately describe the tails of the distribution. The summary of these

posterior predictive checks for pain, urgency, and GRA is in given in Table III. In general, the model fits well, except at week 0 where subjects tended to have higher pain and urgency than predicted by the model and GRA was forced to be 2 by design.

Figure 6 plots the observed Spearman correlation and 95% confidence interval vs. the predicted Spearman correlation and associated 95% posterior predictive interval. Each pair is identified by the first letters of the two variables with I=Wisconsin Total, P=pain, U=Urgency, G=GRA, and C=24-hour frequency. This plot indicates that our model reproduces the observed correlation among the variables fairly well except possibly for correlations involving 24-hour frequency. This is not entirely surprising given that this variable is only measured at 4 of the 5 time points. Twenty-four-hour frequency also has small loadings in both classes (Table I) indicating that it is not contributing substantially to the overall latent severity score.

6. Simulations

In this section we present results of a simulation study designed to assess the performance of the estimation algorithm. Data were simulated under the two-class model in Section 2 using the values in table IV. Simulation parameters were loosely based on the results of the data application. To ease convergence with a modest sample size, variance parameters were chosen to be small and some ordinal outcomes were assumed to be more evenly distributed than in the real data. One hundred data sets were simulated, and each simulated data set had 265 subjects measured at the five times from the data application. One- and two-class models were run using one chain of the MCMC algorithm for 10,000 iterations. Starting values were chosen as described above. The first 5,000 iterations of the MCMC chain were discarded in calculating posterior quantities. We evaluated the performance of our algorithm in several ways. First, we compared DIC values for the one- and two-class models fit to our data and calculated the percentage of time when the two-class model was correctly identified as the superior choice. Second, from the results of the two-class model, we assigned subjects to their most likely classes and assessed the accuracy of this calculation using the Kappa statistic and percent correct classification. Third, within the responder class, we determine whether the lower bound of the pointwise 95% credible interval for the difference in mean curves between treatment and placebo arms ever went above 0, indicating a significant treatment effect in that class. Finally, we calculated the relative bias and nominal 95% coverage of each parameter when the correct model assumed.

When comparing the one-class and two-class models for each data set, the DIC correctly identified the two-class model as the preferable model 82% of the time. Correct classification of subjects was high with average Kappa for agreement of true and estimated class assignments of 0.92 (range of 0.83 to 0.98) and average percent agreement of 96% (range of 91% to 99%). A significant treatment effect was identified in the responder class 100% of the time. The significant treatment effect appeared, on average, at week 7 and not later than week 16.

Figure 7 shows the relative bias of the posterior median for 100 simulated data sets. Most parameters have small relative bias. Notable exceptions are the coefficients for treatment

and treatment by time² interaction terms and the intercept for pain in class 1, the non-responder class. This may be due to the lack of slope or treatment by time effects in this class coupled with fitting a quadratic curve to only 5 points in time.

Figures 8 shows the nominal 95% coverage rates based on the 95% credible intervals. Poor coverage for some of the intercept parameters, particularly in class 2 are of concern. Usually this is related to the presence of bias for the same parameters, however the poor coverage for the intercept for the continuous variable in class one was surprising. This may indicate the need to run longer chains for some data sets. It may also be a function of the relatively small sample size and number of time points for these models.

7. Discussion

In this paper we have presented a Bayesian model which groups subjects based on longitudinal measurements of multiple outcomes summarized through the trajectory of a latent symptom score. Our model accounts for correlation among outcomes at each time point and across time through the use of extended factor analytic and mixed effects framework. We allow for any combination of continuous, binary, ordinal, or count variables observed at any number of time points. Our MCMC algorithm incorporates some non-standard features which help accelerate convergence of the posterior distribution for the threshold parameters and which maintain interpretability of the factor loadings conditional on the latent variable and random effects.

Using this model, we identified a class of subjects where symptoms improved over time and where treatment was effective in substantially reducing symptom severity as reflected by the latent symptom score. Our responder rate of 43% was substantially higher than that of the 16% reported in the primary trial analysis paper [18] suggesting that improvement over time in a complex disease may be better captured by a broader consideration of multiple symptoms measured over time rather than by a single item measured at one time. Our model allows us to simultaneously estimate predictors of class membership and it would be of interest to know what baseline factors might predict whether a subject would respond to treatment to aid in treatment of IC or in design of future clinical trials. We attempted to identify predictors of class membership for inclusion in our class probability model, but were unable to identify any demographic or baseline clinical variables that were associated with class membership. In particular, randomization assignment was not associated with class membership with 43% of placebo patients and 46% of BCG patients being in the responder class. Examination of symptom specific curves revealed that responders may have slightly higher pain and urgency and lower 24-hour frequency at baseline. Further exploration of baseline clinical variables is warranted for identification of potential predictors of symptom improvement.

Our model allows for missing data under the assumption that they are missing at random. There is not much missing data in this IC trial and making this assumption is not problematic. There are however, many cases where the missing data mechanism is not random and ignoring the missing data may lead to incorrect inference. Additional future research may account for such informative missing data.

Our simulation studies indicate that the model performs well in identifying the correct number of classes, in correctly classifying subjects, and in identifying significant treatment effects in the responder class. Bias and coverage were acceptable for most parameters, however, there is some concern with the coverage and bias properties for some of the intercept parameters and parameters in the latent variable model when the true values are zero. Our simulations demonstrate that with 5 repeated measurements we could identify a quadratic curve with a random intercept. The degree of complexity available for the mixed effects model for the latent variable will be determined by the number of repeated measurements available on each subject. While we used a moderate sample size for each simulated data set ($n=265$), performance of the algorithm is no doubt related to sample size, the number of latent classes, variability in the outcomes, and the number of repeated measurements. In a simulation study of univariate growth curve mixture models, Tolvanen [23] suggested sample sizes of greater than 50 for reliable identification of latent classes. The use of informative priors may be beneficial in the small sample setting if there is strong evidence of class characteristics a priori. Further simulation studies are needed to better evaluate the performance of our algorithm under challenging scenarios.

Though we have used some nonstandard features to speed converge of the the threshold parameters, other parameters do not converge quickly and the necessity of Metropolis steps when count, binary, or ordinal outcomes are included slows the algorithm considerably. Further modifications of the algorithm could be pursued to address these issues.

Acknowledgments

Contract/grant sponsor: National Institute for Diabetes and Digestive and Kidney Diseases; contract/grant number: R01-DK59601

APPENDIX

Details of the Gibbs Sampler for the multivariate model with mixed outcomes

I.1. Intercepts and factor loadings

For normally distributed outcomes, the full conditional distribution is

$$\begin{aligned}
 p(\beta_{mk}|\cdot) &\propto \prod_{i=1}^N p(\mathbf{y}_{im}|\alpha_{mk}, \lambda_{mk}, \mathbf{z}_i, b_i, C_i=k, \sigma_m^2)^{c_{ik}} p(\alpha_{mk}) p(\lambda_{mk}) \\
 &\propto \exp \left(-\frac{1}{2} \sigma_m^{-2} \sum_{i=1}^N c_{ik} (\mathbf{y}_{im}^* - \mathbf{z}_i^* \beta_{mk})^T (\mathbf{y}_{im}^* - \mathbf{z}_i^* \beta_{mk}) \right. \\
 &\quad \left. - \frac{1}{2} (\beta_{mk} - \beta_m^*)^T \Theta_{\beta_m}^{-1} (\beta_{mk} - \beta_m^*) \right) \\
 &\propto \exp \left(-\frac{1}{2} \left(\beta_{mk}^T \sigma_m^{-2} \sum_{i=1}^N c_{ik} \mathbf{z}_i^* \mathbf{z}_i^{*T} \beta_{mk} - 2 \beta_{mk}^T \sigma_m^{-2} \sum_{i=1}^N c_{ik} \mathbf{y}_i^{*T} \mathbf{z}_i^* \right. \right. \\
 &\quad \left. \left. + \beta_{mk}^T \Theta_{\beta_m}^{-1} \beta_{mk} - 2 \beta_{mk}^T \Theta_{\beta_m}^{-1} \beta_m^* \right) \right)
 \end{aligned} \tag{19}$$

where $\beta_{mk} = (\alpha_{mk}, \lambda_{mk})^T$. This implies $\beta_{mk} \sim N_2(\mu_{\beta_k}, \Sigma_{\beta_k})$ where

$$\Sigma_{\beta_k} = \left(\sum_{i=1}^N c_{ik} z_i^{*T} z_i^* / \sigma_m^2 + \Theta_{\beta_m}^{-1} \right)^{-1} \text{ and } \mu_{\beta_k} = \Sigma_{\beta_k} * \left(\sum_{i=1}^N c_{ik} z_i^{*T} \tilde{y}_{im} / \sigma_m^2 + \Theta_{\beta_m}^{-1} * \beta_m^* \right).$$

Here $z_i^* = [1 | z_i]$, $\tilde{y}_{im} = y_{im} - b_{im}$, $\beta_m^* = [\alpha_m^*, \lambda_m^*]^T$, and $\Theta_{\beta_m} = \text{diag}(\xi_{\alpha_m}, \xi_{\lambda_m})$.

For ordinal and binary outcomes, y_{im} is replaced by y_{im}^* , the vector of unobserved continuous variables which are sampled as described below. In addition, σ_m^2 is replaced with a vector of subject-and observation-specific variance parameters which follow a Kolmogorov-Smirnov distribution as described in (5) above. Thus,

$$\Sigma_{\beta_k} = \left(\sum_{i=1}^N c_{ik} z_i^{*T} \text{diag}(\sigma_{i1m}^{-2}, \dots, \sigma_{inim}^{-2}) z_i^* + \Theta_{\beta_m}^{-1} \right)^{-1} \text{ and}$$

$$\mu_{\beta_k} = \Sigma_{\beta_k} * \left(\sum_{i=1}^N c_{ik} z_i^{*T} \text{diag}(\sigma_{i1m}^{-2}, \dots, \sigma_{inim}^{-2}) \tilde{y}_{im} + \Theta_{\beta_m}^{-1} * \beta_m^* \right) \text{ where } \tilde{y}_{im} = y_{im}^* - b_{im} \text{ and}$$

the remaining quantities are as defined for the continuous outcomes.

For outcomes following a Poisson distribution,

$$p(\beta_{mk} | \cdot) \propto \prod_{i=1}^N \prod_{j=1}^{n_i} p(y_{ijm} | \alpha_{mk}, \lambda_{mk}, z_i, b_i, C_i = k)^{c_{ik}} p(\alpha_{mk}) p(\lambda_{mk})$$

$$\propto \exp \left(\sum_{i=1}^N \sum_{j=1}^{n_i} \left[c_{ik} y_{ijm} (z_{ij}^{*T} \beta_{mk} + b_{im}) - c_{ik} \exp(z_{ij}^{*T} \beta_{mk} + b_{im}) \right] \right) \quad (20)$$

$$- \frac{1}{2} (\beta_{mk} - \beta_m^*)^T \Theta_{\beta_m}^{-1} (\beta_{mk} - \beta_m^*)$$

where $\beta_m^* = [\alpha_m^*, \lambda_m^*]^T$, and $\Theta_{\beta_m} = \text{diag}(\xi_{\alpha_m}, \xi_{\lambda_m})$.

The full conditional distribution is not available in closed form and so draws proceed via a Metropolis algorithm where the proposal distribution is Multivariate- $t_3(\beta_{mk}^{t-1}, R * V(\beta_{mk}^*))$. β_{mk}^{t-1} is the current value of β_{mk} and $R * V(\beta_{mk}^*) = \hat{V}_k$ is R times the inverse of the Hessian at the mode of the full conditional distribution found using a Newton-Raphson routine. R inflates the variance and is chosen to achieve an acceptance rate of 15–20%. Each complete iteration in the Gibbs chain includes 10 iterations of the Metropolis algorithm so that it is likely that at least one new β_{mk} is accepted.

I.2. Threshold parameters

For ordinal outcomes, the full conditional distribution of the threshold parameters is

$$\begin{aligned}
p(\alpha_{mk}^{(1)}, \dots, \alpha_{mk}^{(D-1)} | \cdot) &\propto \prod_{i,j: C_i=k, y_{ijm}=1} \Phi\left(\frac{\alpha_{mk}^{(1)} - \mu_{ijmk}}{\sigma_{ijm}}\right) \\
&\dots \times \prod_{i,j: C_i=k, y_{ijm}=d} \left[\Phi\left(\frac{\alpha_{mk}^{(d-1)} - \mu_{ijmk}}{\sigma_{ijm}}\right) - \Phi\left(\frac{\alpha_{mk}^{(d)} - \mu_{ijmk}}{\sigma_{ijm}}\right) \right] \\
&\dots \times \prod_{i,j: C_i=k, y_{ijm}=D_m} \Phi\left(\frac{-\mu_{ijmk}}{\sigma_{ijm}}\right) \\
&\times \prod_{d=1}^{D-1} \frac{1}{\sqrt{2\pi*100}} \exp\left(\frac{-\left(\alpha_{mk}^{(d)} - \alpha_{mk}^{(d)*}\right)^2}{2*100}\right) I\left(\alpha_{mk}^{(1)} < \alpha_{mk}^{(2)} < \dots < \alpha_{mk}^{(D-1)}\right)
\end{aligned} \tag{21}$$

where $\mu_{ijmk} = \alpha_{mk} + \lambda_{mk} * z_{ij} + b_{im}$ and Φ is the CDF of a standard normal distribution.

To enhance the mixing properties of the chains for the threshold parameters, we employ the Metropolis-Hastings algorithm of [16] to draw from the full conditional distribution. In brief, we sequentially draw new values for the threshold parameters from a normal proposal distribution $N(\alpha_{mk,t-1}^{(d)}, \sigma_\alpha^2)$ centered at the value of the threshold parameter from the last iteration ($\alpha_{mk,t-1}^{(d)}$) and truncated to $(\alpha_{mk,new}^{(d-1)}, \alpha_{mk,t-1}^{(d+1)})$. The entire vector is accepted with probability

$$\begin{aligned}
R &= \prod_{d=1}^{D-1} \frac{\Phi\left(\left(\alpha_{mk,t-1}^{(d+1)} - \alpha_{mk,t-1}^{(d)}\right)/\sigma_\alpha\right) - \Phi\left(\left(\alpha_{mk,new}^{(d-1)} - \alpha_{mk,t-1}^{(d)}\right)/\sigma_\alpha\right)}{\Phi\left(\left(\alpha_{mk,new}^{(d+1)} - \alpha_{mk,new}^{(d)}\right)/\sigma_\alpha\right) - \Phi\left(\left(\alpha_{mk,t-1}^{(d-1)} - \alpha_{mk,new}^{(d)}\right)/\sigma_\alpha\right)} \\
&\times \prod_{i=1}^N \left[\prod_{j=1}^{J_i} \frac{\Phi\left(\left(\alpha_{mk,new}^{(y_{ijm})} - \mu_{ijmk}\right)/\sigma_{ijm}\right) - \Phi\left(\left(\alpha_{mk,new}^{(y_{ijm}-1)} - \mu_{ijmk}\right)/\sigma_{ijm}\right)}{\Phi\left(\left(\alpha_{mk,t-1}^{(y_{ijm})} - \mu_{ijmk}\right)/\sigma_{ijm}\right) - \Phi\left(\left(\alpha_{mk,t-1}^{(y_{ijm}-1)} - \mu_{ijmk}\right)/\sigma_{ijm}\right)} \right]^{c_{ik}} \\
&\times \prod_{d=1}^{D-1} \frac{\exp\left(-\left(\alpha_{mk,new}^{(d)} - \alpha_{mk}^{(d)*}\right)^2/(2*100)\right)}{\exp\left(-\left(\alpha_{mk,t-1}^{(d)} - \alpha_{mk}^{(d)*}\right)^2/(2*100)\right)}.
\end{aligned} \tag{22}$$

The variance of the proposal distribution, σ_α^2 , is chosen to achieve an acceptance rate of approximately 20% and 5 iterations of the Metropolis algorithm are run within each Gibbs iteration.

For binary and ordinal outcomes, we draw the unobserved underlying continuous variable y_{ijm}^* from its full conditional distribution

$$y_{ijm}^* | \cdot \sim N(\alpha_{mk} + \lambda_{mk} z_{ij} + b_{im}, \sigma_{ijm}^2) \tag{23}$$

truncated to $(\alpha_{mk}^{(y_{ijm}-1)}, \alpha_{mk}^{(y_{ijm})})$ (for ordinal outcomes) or to $(-\infty, 0)$ when $y_{ijm} = 0$ or $(0, \infty)$ when $y_{ijm} = 1$ (for binary outcomes).

I.3. Parameters from the latent variable model: $\mathbf{z}$ and γ

Samples from $p(z_{ij}|\cdot)$ are also obtained using a Metropolis algorithm when count outcomes are present.

The full conditional distribution is

$$\begin{aligned}
 p(z_{ij}|\cdot) &\propto \prod_{k=1}^K \left[\prod_{j=1}^{n_i} \left[\prod_{m=1}^{m_1} p(y_{ijm}|\alpha_{mk}, \lambda_{mk}, z_{ij}, b_i, C_i=k, \sigma_m^2) \right. \right. \\
 &\quad \times \prod_{m=m_1+1}^{m_1+m_2+m_3} p(y_{ijm}^*|\alpha_{mk}, \lambda_{mk}, z_{ij}, b_i, C_i=k, \sigma_{ijm}^2) \\
 &\quad \times \left. \left. \prod_{m=m_1+m_2+m_3+1}^M p(y_{ijm}|\alpha_{mk}, \lambda_{mk}, z_{ij}, b_i, C_i=k) p(z_{ij}|\gamma_k, \mathbf{X}_{ij}^{(Z)}, \mathbf{a}_i, C_i=k) \right] \right]^{c_{ik}} \\
 &\propto \exp \left(\sum_{k=1}^K c_{ik} \left[\sum_{m=1}^{m_1} -\frac{1}{2} \sigma_m^{-2} (y_{ijm} - \alpha_{mk} - b_{im} - \lambda_{mk} z_{ij})^2 \right. \right. \\
 &\quad + \sum_{m=m_1+1}^{m_1+m_2+m_3} -\frac{1}{2} \sigma_{ijm}^{-2} (y_{ijm}^* - \alpha_{mk} - b_{im} - \lambda_{mk} z_{ij})^2 \\
 &\quad + \sum_{m=m_1+m_2+m_3+1}^M y_{ijm} (\alpha_{mk} + \lambda_{mk} z_{ij} + b_{im}) \\
 &\quad \left. \left. - \exp(\alpha_{mk} + \lambda_{mk} z_{ij} + b_{im}) - \frac{1}{2} (z_{ij} - X_{ij}^{(Z)T} (\gamma_k + \mathbf{a}_i))^2 \right] \right).
 \end{aligned}$$

. For draws of the latent variable, a multivariate normal proposal distribution centered at the current value of z_i is used. The variance of the proposal distribution is specified by scaling the inverse of the information matrix found using a Newton-Raphson algorithm.

The full conditional distribution for γ_k is

$$p(\gamma_k|\cdot) \propto \exp \left(-\sum_{i=1}^N c_{ik} \frac{1}{2} (z_i^* - X_i^{(Z)} \gamma_k)^T (z_i^* - X_i^{(Z)} \gamma_k) - \frac{1}{2} (\gamma_k - \gamma^*)^T \Phi^{-1} (\gamma_k - \gamma^*) \right) \quad (24)$$

Thus, γ_k is drawn from $N_p(\boldsymbol{\mu}_{\gamma_k}, \boldsymbol{\Sigma}_{\gamma_k})$ where $\boldsymbol{\Sigma}_{\gamma_k} = \left(\sum_{i=1}^N c_{ik} X_i^{(Z)T} X_i^{(Z)} + \Phi^{-1} \right)^{-1}$ and $\boldsymbol{\mu}_{\gamma_k} = \boldsymbol{\Sigma}_{\gamma_k} * \left(\sum_{i=1}^N c_{ik} X_i^{(Z)T} z_i^* + \Phi^{-1} \gamma^* \right)$. Here $z_i^* = z_i - X_i^Z \mathbf{a}_i$.

I.4. Random effects and variance parameters

For normally distributed outcomes, draw b_{im} from $N(\mu_{b_i}, \sigma_{b_i}^2)$ where $\sigma_{b_i}^2 = (n_i * \sigma_m^{-2} + \tau_m^{-2})^{-1}$ and $\mu_{b_i} = \sigma_{b_i}^2 * \sigma_y^{-2} \tilde{\mathbf{y}}_{mi}^T J_{n_i}$. Here $\tilde{\mathbf{y}}_{mi} = \mathbf{y}_{mi} - \sum_{k=1}^K c_{ik} (\alpha_{mk} + \lambda_{mk} z_i)$ and J_{n_i} is a $n_i \times 1$ vector of 1's.

For binary and ordinal outcomes, draw b_{im} from $N(\mu_{b_i}, \sigma_{b_i}^2)$ where

$$\sigma_{b_i}^2 = \left(\text{diag}(\sigma_{i1m}^2, \dots, \sigma_{inim}^2) + \tau_m^{-2} \right)^{-1} \text{ and } \mu_{b_i} = \sigma_{b_i}^2 * \text{diag}(\sigma_{i1m}^{-2}, \dots, \sigma_{inim}^{-2}) \tilde{\mathbf{y}}_{mi}^T * J_{n_i}. \text{ Here}$$

$$\tilde{\mathbf{y}}_{mi} = \mathbf{y}_{mi}^* - \sum_{k=1}^K c_{ik} (\alpha_{mk} + \lambda_{mk} z_i)$$

For Poisson-distributed outcomes, draw b_{im} using a Metropolis algorithm similar to the draw of β_{mk} .

$$p(b_{im} | \cdot) \propto \prod_{j=1}^{n_i} p(y_{ijm} | \alpha_{mk}, \lambda_{mk}, z_i, b_i, C_i = k)^{c_{ik}} p(b_{im})$$

$$\propto \exp \left(\sum_{j=1}^{n_i} \left[c_{ik} y_{ijm} (z_{ij}^{*T} \beta_{mk} + b_{im}) - c_{ik} \exp(z_{ij}^{*T} \beta_{mk} + b_{im}) \right] - \frac{1}{2} \tau_m^{-2} b_{im}^T b_{im} \right) \quad (25)$$

where $\beta_{mk} = [\alpha_{mk}, \lambda_{mk}]^T$, and $\mathbf{z}_i^* = [\mathbf{1} | \mathbf{z}_i]$.

For all outcomes, draw τ_m^2 from $Inv - \text{gamma}(r_{3m} + N/2, r_{4m} + S_{bm}/2)$ where $S_m = \sum_i b_{im}^2$.

Draw $\mathbf{a}_i$ from $N_q(\boldsymbol{\mu}_{a_i}, \boldsymbol{\Sigma}_{a_i})$ where $\boldsymbol{\Sigma}_{a_i} = (X_i^{zT} X_i + \Psi_a^{-1})^{-1}$ and

$$\boldsymbol{\mu}_{a_i} = \boldsymbol{\Sigma}_{a_i} * X_i^{zT} (\mathbf{z}_i - \sum_{k=1}^K c_{ik} X_i^z \gamma_k). \text{ Next, draw } \Psi_a \text{ from } Inv - \text{Wishart}(N + 1, S_a) \text{ where}$$

$$S_a = \sum_i \mathbf{a}_i \mathbf{a}_i' + I.$$

For normally-distributed outcomes, draw σ_m^2 from

$$Inv - \text{gamma}(r_{1m} + \sum_i n_i / 2, r_{2m} + \sum_i \sum_{j=1}^{n_i} s_{ijm}^2 / 2) \text{ where}$$

$$s_{ijm} = y_{ijm} - \sum_{k=1}^K c_{ik} (\alpha_{mk} + \lambda_{mk} * z_{ij}) - b_{im}. \text{ For ordinal and binary outcomes, the full}$$

conditional distribution of σ_{ijm}^2 is not a standard distribution and so we use rejection sampling via the method of [13] where the proposal density used is the Generalized Inverse

Gaussian(0.5, 1, r^2) distribution with $r^2 = y_{ijm}^* - \alpha_{mC_i} - \lambda_{mC_i} z_{ij} - b_{im}$.

I.5. Class probabilities and class assignments

Draw C_i from *Multinomial*(1, π_i^*) where $\pi_i^* = (\tilde{\pi}_{i1} / \sum_k \tilde{\pi}_{ik}, \dots, \tilde{\pi}_{iK} / \sum_k \tilde{\pi}_{ik})$ and

$$\begin{aligned}
\tilde{\pi}_{ik} = & \pi_{ik} \prod_{j=1}^{n_i} \left[\prod_{m=1}^{m_1} e^{-(y_{ijm} - \alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im})^2 / 2\sigma_m^2} \right. \\
& \times \prod_{m=m_1+1}^{m_1+m_2} \left[\frac{e^{\alpha_{mk}^{(y_{ijm})} - \alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im}}}{1 + e^{\alpha_{mk}^{(y_{ijm})} - \alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im}}} - \frac{e^{\alpha_{mk}^{(y_{ijm}-1)} - \alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im}}}{1 + e^{\alpha_{mk}^{(y_{ijm}-1)} - \alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im}}} \right] \\
& \times \prod_{m=m_1+m_2+1}^{m_1+m_2+m_3} \frac{e^{(1-y_{ijm}) * (-\alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im})}}{1 + e^{-\alpha_{mk} - \lambda_{mk} * z_{ij} - b_{im}}} \\
& \times \left. \prod_{m=m_1+m_2+m_3+1}^M e^{y_{ijm}(\alpha_{mk} + \lambda_{mk} z_{ij} + b_{im}) - \exp(\alpha_{mk} + \lambda_{mk} z_{ij} + b_{im})} \right] e^{-\left(z_{ij} - X_{ij}^{(Z)}(\gamma_k + a_i)\right)^2 / 2} \quad (26)
\end{aligned}$$

When the class probability model contains only an intercept, we draw the class probabilities directly from the full conditional distribution

$$\pi_1, \dots, \pi_K \sim \text{Dirichlet}(\omega_1, \dots, \omega_K) \quad (27)$$

$$\text{where } \omega_k = \sum_{i=1}^N c_{ik} + \omega^*$$

When the probability of class membership is a function of covariates, we draw δ_k via a Metropolis algorithm where

$$p(\delta_k | \cdot) \propto \prod_{i=1}^N \prod_{k=1}^K \left(\exp(X_i^{(C)} \delta_k) / \sum_k \exp(X_i^{(C)} \delta_k) \right)^{c_{ik}} \exp\left(-\frac{2}{9} \delta_k^T \delta_k\right). \quad (28)$$

REFERENCES

1. Leiby BE, Sammel MD, Ten Have TR, Lynch KG. Identification of treatment responders in an interstitial cystitis clinical trial using a bayesian multivariate growth curve latent class model. *Applied Statistics*. 2009; 58:505–524. [PubMed: 21637724]
2. Roy J, Lin X. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*. 2000; 56(4):1047–1054. [PubMed: 11129460]
3. Muthén B. A structural probit model with latent variables. *Journal of the American Statistical Association*. 1979; 74:807–811.
4. Muthén B. Latent variable structural equation modeling with categorical data. *Journal of Econometrics*. 1983; 22:43–65.
5. Muthén B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984; 49:115–132.
6. Arminger G, Küsters U. Latent trait models with indicators of mixed measurement level. *Latent Trait and Latent Class Models*. 1988:51–73.
7. Dunson DB. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B, Methodological*. 2000; 62(2):355–366.
8. Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B, Methodological*. 1997; 59:667–678.
9. Moustaki I, Knott M. Generalized latent trait models. *Psychometrika*. 2000; 65(3):391–411.
10. Dunson DB. Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*. 2003; 98:555–563.

11. Miglioretti DL. Latent transition regression for mixed outcomes. *Biometrics*. 2003; 59:710–720. [PubMed: 14601773]
12. Chen MH, Dey D. Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhya, Series A*. 1998; 60:322–343.
13. Holmes CC, Held L. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*. 2006; 1:145–168.
14. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1(3):515–533.
15. Garrett ES, Zeger SL. Latent class model diagnosis. *Biometrics*. 2000; 56:1055–1067. [PubMed: 11129461]
16. Cowles M. Accelerating monte carlo markov chain convergence for cumulativelink generalized linear models. *Statistics and Computing*. 1996; 6:101–111.
17. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992; 7:457–472.
18. Mayer R, Probert KJ, Peters KM, Payne CK, Zhang Y, Burks D, Culkin DJ, Diokno A, Hanno P, Landis JR, et al. A randomized controlled trial of intravesical bacillus calmette-guerin for treatment refractory interstitial cystitis. *Journal of Urology*. 2005; 173(4):1186–1191. [PubMed: 15758738]
19. Goin J, Olaleye D, Peter K, Steinert B, Habicht K, Wynant G. Psychometric analysis of the university of wisconsin interstitial cystitis scale: implication for use in randomized clinical trials. *Journal of Urology*. 1998; 159:1085. [PubMed: 9474236]
20. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, Methodological*. 2002; 64(4):583–616.
21. Celeux G, Forbes F, Robert C, Titterton D. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006; 1(4):651–674.
22. Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*. 1996; 6:733–760.
23. Tolvanen. Report 111. University of Jyväskylä Department of Mathematics and Statistics, PO Box 35 (MaD) FI-40014 University of Jyväskylä Finland; 2007. Latent growth mixture modeling: A simulation study.

Figure 1. Results from the one-class model. (Left) Mean latent symptom curves by treatment group. Solid Line=Placebo, Dotted Line=BCG. (Right) Mean difference in treatment and placebo arms with 95 % pointwise credible intervals. Positive values indicate that the treatment arm has less severe symptoms.

Figure 2. Results from the two-class model. Mean latent symptom curves by treatment group for “Non-responders” (Left) and “Responder” (Right). Solid line=Placebo, Dotted line=BCG.

Figure 3. Results from the two-class model. Mean difference in treatment and placebo arms with 95 % pointwise credible intervals for “Non-responders” (Left) and “Responder” (Right). Positive values indicate that the treatment arm has less severe symptoms.

Figure 4. Results from the two-class model. Mean symptom curves by class (black=responder, gray=nonsreponder) and treatment arm (solid=placebo, dashed=treatment).

Figure 5.

QQ-plots comparing the observed distribution with the posterior predictive distribution at week 0 for the Wisconsin Total Score (left) and 24-hour frequency (right).

Figure 6.
Posterior predictive correlation vs. observed correlation

Figure 7.

Boxplots of Relative Bias for 100 simulated data sets. Clockwise from lower left: Variance parameters, Class 1 (Non-responder) parameters, Class 2 (Responder) parameters.

Figure 8.

Nominal 95% coverage for 100 data sets. Clockwise from bottom left: Variance Parameters, Class 1 (Non-responder) parameters, Class 2 (Responder) parameters,

Table 1

Median posterior estimates and 95% credible intervals for the one- and two-class multivariate growth curve latent class models. Outcome-specific parameters.

Parameter		One-Class Model	Two-Class Model	
			Class 1 Non-responders	Class 2 Responders
Pr(C=k)		-	0.57 (0.48,0.67)	0.43 (0.33,0.52)
Wisconsin Total Score	α	27 (25,28)	31 (29,32)	24 (22,26)
	λ	4.7 (4.4,5.0)	2.0 (1.5,2.7)	5.7 (5.1, 6.2)
	σ^2	19.1 (16.7,21.5)	16.2 (14.3,18.5)	
	τ^2	20 (15,26)	17.9 (13.4,23.7)	
24-hour frequency	α	2.8 (2.7,2.8)	2.9 (2.8,3.0)	2.6 (2.5,2.7)
	λ	0.11 (0.09,0.13)	0.09 (0.07,0.12)	0.09 (0.07,0.11)
	τ^2	0.13 (0.10,0.16)	0.12 (0.10,0.14)	
Pain	α	-0.9 (-2.1,0.3)	-0.1 (-1.6,1.5)	-1.8 (-3.3,-.5)
	λ	3.6 (2.9,4.7)	3.4 (2.4,4.7)	4.0 (3.1,5.2)
	$\alpha^{(1)}$	-6.4 (-8.0,-5.2)	-6.4 (-8.6,-4.9)	-7.2 (-9.2,-5.6)
	τ^2	4.7 (2.3,9.0)	5.0 (2.3,9.3)	
Urgency	α	-0.2 (-1.2,0.9)	0.5 (-0.7,1.8)	-0.7 (-2.0,0.5)
	λ	3.0 (2.6,3.7)	2.6 (1.9,3.3)	3.2 (2.6,3.9)
	$\alpha^{(1)}$	-5.9 (-7.0,-5.1)	-6.2 (-7.7,-4.9)	-6.0 (-7.2,-4.9)
	τ^2	5.2 (3.2,8.9)	5.6 (3.4,8.8)	
GRA	α	6.7 (4.7,9.4)	7.2 (5.5,10.0)	5.7 (3.8,8.3)
	λ	3.4 (2.5,4.5)	2.4 (1.4,3.9)	3.4 (2.5,4.8)
	τ^2	9.6 (3.8,20.9)	8.5 (3.9,19.8)	

Table 2

Median posterior estimates and 95% credible intervals for the two-class multivariate growth curve latent class model. Growth Curve parameters.

Parameter	One-Class Model	Two-Class Model	
		Class 1 Non-responders	Class 2 Responders
time/10	−0.19 (−0.27,−0.12)	0.03 (−0.11,0.16)	−0.41 (−0.57,−0.29)
time ² /1000	2.4 (1.6,3.2)	2.7 (1.5,3.9)	2.6 (1.5,3.7)
arm	−0.4 (−0.8,−0.1)	0 (−0.5,0.5)	−.9 (−1.5,−0.4)
arm*time/10	−0.13 (−0.24,−0.02)	0 (−0.2,0.2)	−.3 (−0.5,−.1)
arm*time ² /1000	0.5 (−0.5,1.6)	−0.3 (−1.9,1.4)	1.5 (−0.01,3.3)
age	−0.03 (−0.09,0.04)	−0.10 (−0.18,−0.02)	0.03 (−0.05,0.12)
log ₁₀ duration	−0.20 (−0.56,0.17)	0.27 (−0.23,0.78)	−0.92 (−1.60,−0.46)
Ψ_a	1.4 (1.1,1.8)	0.8 (0.5,1.2)	

Table 3

Posterior predictive check of Pain, Urgency, and GRA

Week	Pain		Urgency		GRA	
	Mean χ^2	$Pr(\chi^2 > 5.99)$	Mean χ^2	$Pr(\chi^2 > 5.99)$	Mean χ^2	$Pr(\chi^2 > 5.99)$
0	19.8	0.99	15.93	0.72	10.94	0.99
8	4.45	0.27	5.13	0.33	2.34	0.21
18	5.62	0.36	3.03	0.13	1.88	0.16
26	3.99	0.24	3.13	0.14	2.44	0.25
34	4.65	0.28	4.40	0.27	1.80	0.16

Table 4

True parameter values assumed for the Simulations Studies

Parameter		Class 1	Class 2
Pr(C=k)		0.55	0.45
Wisconsin Total Score	α	31	27
	λ	1.5	3.7
	σ^2	2	
	τ^2	2	
24-hour frequency	α	2.9	2.6
	λ	0.1	0.09
	τ^2	0.12	
Pain	α	-0.62	-0.61
	λ	1.3	2.0
	$\alpha^{(1)}$	-1.71	-2.24
	τ^2	0.25	
Urgency	α	-0.71	-0.93
	λ	1.5	2.0
	$\alpha^{(1)}$	-1.95	-2.56
	τ^2	0.25	
GRA	α	-0.49	-0.3
	λ	1.0	2.0
	τ^2	0.25	
Growth Curve Parameters	time/10	0.1	-0.4
	time ² /1000	2.7	2.66
	arm	0.1	-0.84
	arm*time/10	-0.05	-0.25
	arm*time ² /1000	-0.24	1.43
	Ψ_a	0.4	