Biometrics 59, 710–720
September 2003

Latent Transition Regression for Mixed Outcomes

Diana L. Miglioretti

Group Health Cooperative, Center for Health Studies, 1730 Minor Ave. Suite 1600, Seattle, Washington 98101, U.S.A.
email: miglioretti.d@ghc.org

Summary. Health status is a complex outcome, often characterized by multiple measures. When assessing changes in health status over time, multiple measures are typically collected longitudinally. Analytic challenges posed by these multivariate longitudinal data are further complicated when the outcomes are combinations of continuous, categorical, and count data. To address these challenges, we propose a fully Bayesian latent transition regression approach for jointly analyzing a mixture of longitudinal outcomes from any distribution. Health status is assumed to be a categorical latent variable, and the multiple outcomes are treated as surrogate measures of the latent health state, observed with error. Using this approach, both baseline latent health state prevalences and the probabilities of transitioning between the health states over time are modeled as functions of covariates. The observed outcomes are related to the latent health states through regression models that include subject-specific effects to account for residual correlation among repeated measures over time, and covariate effects to account for differential measurement of the latent health states. We illustrate our approach with data from a longitudinal study of back pain.

Key words: Hierarchical model; Latent class analysis; Latent traits; Latent transition analysis; Longitudinal data; Markov chain Monte Carlo; Multivariate response.

1. Introduction

Multivariate outcomes of mixed types occur frequently in many research areas including birth defects in teratology (Sammel, Ryan, and Legler, 1997), cognitive and functional ability in Alzheimer's research (Gray and Brookmeyer, 2000), and pain in public health research (Von Korff et al., 1992; Sammel and Landis, 1998). Typically, multiple measures are collected when the outcome of interest is difficult to quantify. For example, when studying chronic back pain, continuous and categorical measures of pain intensity, persistence, and interference with daily activities are collected for lack of a single suitable measure of pain severity. Scientific interest is in prognostic factors for the onset and continuation of severe pain, characterized by a combination of these measures.

The analysis of multivariate longitudinal outcomes is especially challenging when the outcomes are of mixed type. Longitudinal regression approaches for multivariate data analysis may be divided into three broad groups. Perhaps the simplest approach is to summarize the multiple outcomes into a single score, such as an average or sum, and analyze this summary measure in a univariate, longitudinal regression model. However, summary measures are often chosen subjectively and may result in a significant loss of information. In addition, it may not be clear how to summarize mixtures of continuous and discrete data.

A second approach is to regress each outcome on covariates of interest and summarize the results. The regression coefficients may be combined into a single estimate using a global test for a common treatment effect (e.g., O'Brien, 1984; Pocock, Geller, and Tsiatis, 1987; Legler, Lefkopoulou, and Ryan, 1995; Sammel, Lin, and Ryan, 1999). When outcomes are of different types, this approach is complicated by the fact that regression coefficients have different interpretations. A creative solution to this problem was offered by Gray and Brookmeyer (2000), who proposed a treatment effect measure that is a transformation of the time scale and may be combined across continuous, discrete, and time-to-event data.

A third approach common in the social sciences, and of increasing interest in the health sciences, is to summarize and analyze multiple outcomes in a single model by treating the multiple measures as surrogates of an underlying latent variable that represents true health status, directly regressing the latent variable on covariates. Typically, longitudinal latent variable regression models have been limited to multivariate outcomes of a single type (*continuous outcomes:* Sammel and Ryan, 1996; Lin et al., 2000; Roy and Lin, 2000; *categorical outcomes:* Legler and Ryan, 1997; Reboussin, Liang, and Reboussin, 1999; Humphreys and Janson, 2000; *Poisson outcomes:* Roeder, Lynch, and Nagin, 1999). Sammel et al. (1997) proposed a latent variable model for cross-sectional mixed outcomes using a generalized linear mixed model with continuous latent variables, allowing covariate effects on both the outcomes and the latent variables. Muthén and Shedden (1999) proposed a general latent variable modeling framework that incorporates both continuous and categorical outcomes and associates separate latent variables for the outcomes of each type.

This article focuses on the third approach and introduces a latent transition regression model for mixed outcomes. Health status is treated as a categorical latent variable, and both baseline health status prevalences and transition probabilities between these latent health states are modeled as functions of covariates. The observed outcomes are related to the latent health states through regression models that include subject-specific effects to account for residual correlation among repeated measures within subjects over time. They may include covariates to account for *differential measurement* of the latent health states, i.e., differences in the definition of latent health states for subgroups of individuals.

The latent transition regression model for mixed outcomes is introduced in the next section. Model fitting using Bayesian sampling techniques is described in Section 3. An example of an application to chronic back pain is provided in Section 4. The article concludes with a discussion of the proposed methodology, and an Appendix gives computational details.

2. The Latent Transition Regression Model

The latent transition regression model for mixed outcomes includes a *structural model* for the underlying latent health states and a *measurement model* for the observed data, conditional on latent health status. For the structural model, we assume there are J latent health states indicated by $S_{i0}, \ldots, S_{iT_i-1}$ for the ith subject at a set of T_i discrete time points; $i = 1, \ldots, N$. Baseline health state prevalences $p(S_{i0} = j) = \eta_{ij}$ and the probabilities of transitioning between these health states $p(S_{it} = j \mid S_{it-1} = k) = \tau_{it}(j,\ k)$ are modeled as functions of covariates $\mathbf{x}_{it}$ using polytomous regression (Agresti, 1984); $j,\ k = 1, \ldots, J$. A first-order Markov process is assumed here for simplicity:

$$S_{i0} \mid \boldsymbol{\alpha} \sim \text{Multinomial}(\eta_{i1}, \ldots, \eta_{iJ}) \tag{1}$$

$$\log\left(\frac{\eta_{ij}}{\eta_{iJ}}\right) = \boldsymbol{\alpha}_j' \mathbf{x}_{i0}$$

$$S_{it} \mid (S_{it-1} = k, \beta) \sim \text{Multinomial}\{\tau_{it}(1,k), \ldots, \tau_{it}(J,k)\}$$

$$\log\left\{\frac{\tau_{it}(j,k)}{\tau_{it}(J,k)}\right\} = \boldsymbol{\beta}_{tjk}' \mathbf{x}_{it}, \tag{2}$$

where $\boldsymbol{\alpha}_J$ and $\boldsymbol{\beta}_{tJk} = 0$. Equation (2) is a transition model since the regression coefficients $\boldsymbol{\beta}_{tjk}$ depend on k, the latent class membership at the previous time point. It is often reasonable and useful to make the simplifying assumption that $\boldsymbol{\beta}_{tjk}$ does not depend on time t.

Let $y_{i1t}, \ldots, y_{iMt}$ be the M observed mixed outcomes for the ith individual at time t. For the measurement model, the mth outcome may follow any parametric distribution $F(\boldsymbol{\theta}_{imtj},\ \boldsymbol{\zeta}_{mj})$. We partition the parameters into two sets $\boldsymbol{\theta}_{imt}$ and $\boldsymbol{\zeta}_{mj}$. Both sets may depend on latent health status $S_{it} = j$. The $C \times 1$ parameter vector $\boldsymbol{\theta}_{imtj}$ may be additionally modeled as a function of subject-specific effects $\mathbf{u}_{im}$ and covariates $\mathbf{z}_{it}$ that may include the same or different covariates as $\mathbf{x}_{it}$:

$$y_{imt} \mid (S_{it} = j) \sim F(\boldsymbol{\theta}_{imtj}, \boldsymbol{\zeta}_{mj})$$

$$g_m(\theta_{imtjc} \mid S_{it} = j) = \gamma_{mjc} + u_{imc} + \boldsymbol{\delta}_{mc}' \mathbf{z}_{it}$$

$$u_{imc} \sim N(0, 1/\xi_m),\ c = 1, \ldots, C,$$

where $g(\cdot)$ is some link function (McCullagh and Nelder, 1989). For example, if $y_{imt} \sim \text{Normal}(\theta_{imt},\ \zeta_{mj})$, the mean θ_{imt} may be modeled as a linear function of health status $S_{it} = j$, subject-specific effects u_{im}, and covariates $\mathbf{z}_{it}$, while the variance ζ_{mj} depends on $S_{it} = j$ but is held constant across individuals within classes. Other examples of $F(\boldsymbol{\theta}_{imtj},\ \boldsymbol{\zeta}_{mj})$ are given in Section 4.2.

There are several notable features of the measurement model. First, the intercepts and other regression parameters do not depend on time, which constrains the health status definitions to be constant over time. Typically, more meaningful models have stable health status definitions, although there may be applications for which it makes sense to relax this assumption in a constrained way. Second, the subject-specific effects $\mathbf{u}$ do not depend on latent health status. The subject-specific effects are meant to capture a person's tendency to differ from their expected values, independent of their health status. Third, the subject-specific effects $\mathbf{u}_i$ are independent across the observed outcome measures $y_{i1}, \ldots, y_{iM}$ given $S_{i1}, \ldots, S_{iT-1}$. This is a standard and often reasonable assumption of latent variable models. If residual dependence exists, more classes could be fit or correlation among the $\mathbf{u}_i$ may added by assuming $\mathbf{u}_i$ follows a multivariate normal distribution with some covariance matrix. Finally, the covariate effects $\boldsymbol{\delta}_m$ do not depend on health status. This constraint is imposed to ensure identifiability of the covariate effects in the case when both the observed outcomes and the latent health state prevalences and transition probabilities depend on the same covariates (Huang and Bandeen-Roche, 2003, in press). The inclusion of covariates in the regression of the observed outcomes allows for differential measurement, i.e., differences in the definition of a latent health status for subgroups of individuals. However, for our purposes, this is not the main regression of interest. The regression of interest is on the latent health states, defined above in equations (1) and (2).

Given health status S_{it} at time t and the subject-specific effects $\mathbf{u}_i$, the observed outcomes $\mathbf{y}_{it}$ at time t are assumed to be (1) independent of health status at all other time points, (2) independent within and across time points, and (3) independent of the covariates $\mathbf{x}_{it}$ not included in the covariate vector $\mathbf{z}_{it}$. Along with the assumption of independence among individuals, these assumptions give the following complete-data likelihood, conditional on latent health states:

$$p(y \mid \cdot) = \prod_{i=1}^{N} \prod_{t=0}^{T_i-1} \prod_{m=1}^{M} p\{y_{imt} \mid S_{it}, \boldsymbol{\theta}_{imtS_{it}}(\boldsymbol{\gamma}_{S_{it}m}, \mathbf{u}_i, \boldsymbol{\delta}_m), \zeta_{mS_{it}}\}. \tag{3}$$

where $\mid \cdot$ denotes conditioning on all other model parameters. The observed-data likelihood may be obtained by marginalizing over the latent class distributions:

$$p\left(y \mid \boldsymbol{\theta}\left(\boldsymbol{\gamma}, \mathbf{u}, \boldsymbol{\delta}\right), \boldsymbol{\zeta}\right)$$

$$= \prod_{i=1}^{N} \sum_{j_0=1}^{J} \cdots \sum_{j_{T_i-1}=1}^{J} p\left(S_{i0} = j_0\right) \prod_{t=1}^{T_i-1} p\left(S_{it} = j_t \mid S_{it-1} = j_{t-1}\right)$$

$$\times \prod_{t=0}^{T_i-1} \prod_{m=1}^{M} p\{y_{imt} \mid S_{it} = j_t, \boldsymbol{\theta}_{imtj_t}(\boldsymbol{\gamma}_{j_t m}, \mathbf{u}_i, \boldsymbol{\delta}_m), \zeta_{mj_t}\}.$$

3. Model Fitting

We use Markov chain Monte Carlo (MCMC) (Gilks, Richardson, and Spiegelhalter, 1996) to sample from the joint posterior distribution, which is proportional to the product of the prior distributions, the probabilities of the latent health states, and the complete-data likelihood (3):

$$p(\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, u, \boldsymbol{\zeta}, S \mid y) \propto p(\boldsymbol{\alpha})p(\beta)p(\boldsymbol{\gamma})p(\boldsymbol{\zeta}) \times \prod_{i=1}^{N}\prod_{m=1}^{M} p(u_{im} \mid \zeta_m)p(S_{i0} \mid \boldsymbol{\alpha}) \times \prod_{t=1}^{T_i-1} p(S_{it} \mid S_{it-1}, \beta)p(\mathbf{y}_i \mid \mathbf{S}_i, \mathbf{u}_i, \boldsymbol{\gamma}, \delta, \zeta). \tag{4}$$

We present our fitting approach using standard prior distributions, taking $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ to be normally distributed and $\boldsymbol{\zeta}$ to be gamma-distributed with investigator-specified hyperparameters. Each parameter or block of parameters is updated, in turn, by conditioning on all other parameters in the style of Gibbs sampling. The latent health states may be sampled directly from their full conditional distributions:

$$p(S_{i0} = j \mid \cdot) = \frac{p(S_{i0} = j \mid \boldsymbol{\alpha})p(S_{i1} \mid S_{i0} = j, \boldsymbol{\beta})p(\mathbf{y}_{i0} \mid S_{i0} = j, \boldsymbol{\theta}, \boldsymbol{\zeta})}{\sum_{k=1}^{J} p(S_{i0} = k \mid \boldsymbol{\alpha})p(S_{i1} \mid S_{i0} = k)p(\mathbf{y}_{i0} \mid S_{i0} = k, \boldsymbol{\theta}, \boldsymbol{\zeta})}$$

$$p(S_{it} = j \mid \cdot) = \frac{p(S_{it} = j \mid S_{it-1}, \boldsymbol{\beta})p(S_{it+1} \mid S_{it} = j)p(\mathbf{y}_{it} \mid S_{it} = j, \boldsymbol{\theta}, \boldsymbol{\zeta})}{\sum_{k=1}^{J} p(S_{it} = k \mid S_{it-1}, \boldsymbol{\beta})p(S_{it+1} \mid S_{it} = k)p(\mathbf{y}_{it} \mid S_{it} = k, \boldsymbol{\theta}, \boldsymbol{\zeta})}, \quad t = 1, \ldots, T_i - 1.$$

Given the latent health states, the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ may be updated in separate blocks using random walk Metropolis steps (Metropolis et al., 1953). Tuning parameters for updating these blocks may be estimated using the three-simulation strategy described in Raftery and Lewis (1996). The measurement model parameters, $\boldsymbol{\gamma}, \boldsymbol{\delta}$, $\mathbf{u}$, and $\boldsymbol{\zeta}$, may be updated using Metropolis steps or may be directly sampled from their full posterior distributions, depending on their assumed distributions. The subject-specific precisions $\boldsymbol{\zeta}$ may be directly sampled from their full conditional distributions, which are Gamma$(A + N/2, B + 1/2\sum_{i=1}^{N} u_{im}^2)$. Details of the MCMC algorithm are given in the Appendix.

4. Example

The proposed methods are illustrated with data from a longitudinal study of primary care chronic back pain patients conducted at Group Health Cooperative (Von Korff et al., 1992, 1993, 1994; Dionne et al., 1997, 1999). The study consisted of 1213 back pain patients between 18 and 75 years of age who visited their physician due to their back pain at baseline. Fifty-two percent were female. Each patient was given a 30-minute phone interview at baseline, 1, 2, and 5 years. The main question of interest is whether current depression is a prognostic factor for future chronic pain, defined by continued severe pain, adjusting for age, sex, and education. Several studies have shown that psychological distress in general, and depression in particular, predict poor outcomes among back pain patients (Pincus et al., 2002). If depression level is found to be an important prognostic factor in this study, it would support the idea that depression level should be considered when determining the appropriate treatment options for primary care back pain patients.

4.1 *Structural Model*

To answer the questions of interest, it is useful to assume the study population is made of people with no back pain (at follow-up), people with mild-to-moderate pain, and people with severe pain. Therefore, we initially fitted a model with three latent health states. Model diagnostics revealed the presence of a fourth group of patients at baseline with more severe pain than was found at the follow-up points, which is reasonable given that patients were selected because of visiting their physicians for back pain. There was no evidence that this more severe class remained at the follow-up points; therefore, the prevalence was constrained to be zero at these time points for model identifiability. The model would not converge without this constraint. The prevalence of the *no pain* status was constrained to be zero at baseline, because back pain was an inclusion criterion for the study.

Using polytomous regression, we modeled the baseline class prevalences as a function of baseline depression status measured by the Hopkins symptoms checklist (SCL) (Derogatis et al., 1974), adjusting for age, sex, and education. The transition probabilities were modeled as a function of depression score at the previous visit, adjusting for age, sex, and education. Given differences in time intervals between visits, the regression coefficients were allowed to vary over time.

4.2 *Measurement Model*

Five measures of pain severity were included in the analysis: (1) *Pain intensity* measures pain level in the past 3 months. (2) *Pain interference* and (3) *pain impact* measure how much back pain interferes with and impacts daily activities and work. (4) *Number of disability days* is the number of days in the prior six months that the patient was unable to carry out usual activities due to back pain. (5) *Kept from full-time work* indicates whether the subject was kept from full-time work in past year due to back pain. The *no pain* group ($j = 1$) was constrained to have all outcomes equal zero. The model fully estimated the definitions of the other three pain states based on patterns in the data.

Pain intensity, interference, and *impact* are continuous measures scaled to range from 0 to 100. For $j > 1$, we modeled these outcomes using truncated normal (TN) distributions (Johnson and Kotz, 1970) with means that depend on latent pain status and subject-specific effects:

$$y_{imt} \mid (S_{it} = j > 1) \sim \mathrm{TN}\left(\mu_{imj}, \sigma_{mj}^2\right)$$
$$\mu_{imj} = \gamma_{mj} + u_{im},\ m = 1, 2, 3.$$

The truncated normal is a flexible distribution that can describe both floor and ceiling effects, since it can be skewed in either direction (Rutter and Miglioretti, 2003).

We first modeled the *number of disability days* using a Poisson distribution; however, model diagnostics revealed that the Poisson distribution did not adequately fit the data. As an alternative, we categorized the data into the following five groups: 0 days, 1–6 days, 7–14 days, 15–30 days, and >30 days, using a multinomial distribution. For $j > 1$, we modeled the category prevalences as functions of latent pain status and subject-specific effects using polytomous regression:

$$y_{i4t} \mid (S_{it} = j) \sim \text{Multinomial}\,(\pi_{i4j1}, \ldots, \pi_{i4j5})$$
$$\log\left(\frac{\pi_{i4jc}}{\pi_{i4j5}}\right) = \gamma_{4jc} + u_{i4c},$$

where $\gamma_{4j5} = u_{i45} = 0$.

For $j > 1$, we modeled the probability of being *kept from full-time work* using logistic regression. Since a person who is retired or a homemaker may be less likely to answer that they were kept from full-time work, we accounted for differential measurement of this variable by modeling the response probability as a function of being retired or a homemaker ($z_{it} = 1$), in addition to pain status and subject-specific effects. Based on model diagnostics, the probability of being *kept from full-time work* and the effect of being retired/homemaker on this probability was allowed to be different at baseline:

$$y_{i5t} \mid (S_{it} = j) \sim \text{Bernoulli}(\pi_{i5tj})$$
$$\text{logit}(\pi_{i5tj}) = \gamma_{5j} + u_{i5} + \delta_1 I(t = 0) + \delta_2 z_{it} + \delta_3 z_{it} I(t = 0),$$

where $I(\cdot)$ is the indicator function.

The prior distributions for the truncated normal outcome intercepts γ_{1j}, γ_{2j}, and γ_{3j} were taken to be Uniform$(-10, 110)$. Using prior ranges that are slightly larger than the range of the data allows flexibility in the distributional shape, while assuring the mean does not become extremely small or large where the truncated normal distribution is not well behaved (Rutter and Miglioretti, 2003). The prior distributions for the logistic and polytomous regression coefficients $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_4, \boldsymbol{\gamma}_5, \boldsymbol{\delta})$ were taken to be Normal(0, 9/4), which is a noninformative prior that is relatively flat for the inverse-logit of logistic regression parameters and favors equal multinomial probabilities for polytomous regression (Garrett and Zeger, 2000). The precisions σ_{mj}^{-2} for the continuous outcomes $m = 1, 2, 3$ were taken to be Gamma(0.01, 0.01). The subject-specific effects $\mathbf{u}_i$ were assumed to follow Normal(0, ζ_{mc}) distributions and the prior distribution for the precisions ζ_{mc} were taken to be Gamma(0.01, 0.01).

4.3 *Model Fitting*

Details of the MCMC algorithm are given in the Appendix. We first fitted a latent transition model without covariate effects to obtain good starting parameters for the latent pain status definitions. We started the sampler at three different points in the parameter space to ensure the model was converging to the same estimates. We ran the samplers for 5000 iterations and discarded the first 2000 iterations for burn-in. We used the means across the 9000 iterations from the combined samplers as starting values for the model with covariates. For the regression model, we started the regression coefficients at three different sets of random points in the parameter space and verified that the samplers converged to the same estimates. We ran each sampler for 100,000 iterations, discarded the first 40,000 iterations for burn-in, and kept every 20th iteration. The results are based on the remaining 9000 iterations from the three independent samplers.

4.4 *Results*

The latent pain status definitions are displayed in Table 1. We estimated the distributions of the population average parameters by integrating over the random effect distributions using Gauss-Hermite numerical quadrature with 20 points. The first pain status corresponds to people with *no pain* and all values were constrained to be zero. We labeled the second status *moderate pain/low interference* based on the parameter estimates. For this status, the population-average sample mean *pain intensity* is 27, which is moderate. The mean *pain interference* and *pain impact* are 14 and 13, the majority of people (80%) have zero reported *disability days,* and work disability is very low, indicating low interference. The third pain status corresponds to people with *severe pain/moderate interference* with daily activities. *Pain intensity* is high with a mean of 52. *Interference* and *impact* are moderate with means of 45 and 40, respectively. The majority of people (73%) have one or more *disability days;* 20% of people who are not retired or homemakers report being *kept from full-time work* due to back pain in the past year at follow-up visits and 8% report being *kept from full-time work* at baseline. The last class is a *severe pain/high interference* group that is only present at baseline. *Pain intensity, interference,* and *impact* are all high, with means of 69, 71, and 65. The majority of people (81%) reported more than two weeks of *disability days* and 16% of people who are not retired or homemakers reported being *kept from full-time work*. In general, the probability of being *kept from full-time work* due to back pain in the past year is lower at baseline, which is reasonable since it is more likely to be a patient's first episode of back pain at baseline due to the design of the study.

Also shown in Table 1 are the latent pain state prevalences. These were directly estimated from the sampled class memberships by calculating the mean of the percentage of people in each class at each time period. At baseline, the majority of patients (76%) have *severe pain*, 58.6 with *moderate interference* and 17.5% with *high interference*. No patients have *no pain,* by the design of the study. The prevalence of *severe pain* decreased to 46% in year one, 39% in year two, and 36% in year five. The prevalence of *no pain* increased to 11% at year one, to 16% in year two, and up to 22% by year five. The probability of *severe pain with high interference* was constrained to be zero at the follow-up points, for model identifiability (see Section 4.1).

The transition probability matrices, calculated from the sampled class memberships, are displayed in Table 2. The columns correspond to pain status at time t and the rows correspond to time $t - 1$. These matrices differ somewhat over time. For all time points, the probabilities on the diagonal are

Table 1
Latent pain status definitions and prevalences. Constrained parameters are in italics.

	Pain status definitions			
	No pain	Moderate pain low interference	Severe pain mod interference	Severe pain high interference
Truncated normal outcomes		Sample mean (SD)		
(1) Pain intensity	0	27 (11)	52 (11)	69 (10)
(2) Pain interference	0	14 (7)	45 (18)	71 (17)
(3) Pain impact	0	13 (7)	40 (21)	65 (22)
Multinomial outcome				
(4) Number of disability days		Category prevalences		
0 days	1	0.80	0.27	0.03
1 to 6 days	0	0.16	0.26	0.04
7 to 14 days	0	0.03	0.19	0.11
15 to 30 days	0	0.00	0.11	0.26
31–180 days	0	0.00	0.16	0.55
Binary outcomes		Response probabilities		
(5) Kept from FT work				
Baseline:	N/A	0.007	0.078	0.16
for retired/homemakers	N/A	0.001	0.010	0.025
Follow-Up:	0	0.024	0.20	N/A
for retired/homemakers	0	0.017	0.15	N/A
	Pain status prevalences			
	No pain	Moderate pain low interference	Severe pain mod interference	Severe pain high interference
Baseline (N = 1,116)	0	0.24	0.59	0.18
One year (N = 1,022)	0.11	0.43	0.46	0
Two years (N = 897)	0.16	0.45	0.39	0
Five years (N = 701)	0.22	0.42	0.36	0

the largest for each row, indicating that, on average, a person's most likely pain state at each time point is their previous pain state. About 59% of people with *no pain* at one and two years continue to have *no pain* at the following time points. Most of those who get worse progress to *moderate pain/low interference*. The percentage of patients with continued *moderate pain* ranges from 57% to 73%. Among the patients that move out of this pain state, the majority get better. Eighty-three percent of patients with *severe pain/high interference* at baseline have *severe pain* at one year, and between 51% and 72 patients with *severe pain/moderate interference* have continued *severe pain* at the next time point, i.e., *chronic pain.* Only 4% to 7% of these patients improve to the point of having *no pain* at the next time point.

Table 2
Transition probabilities. Rows correspond to time $t-1$. Columns correspond to time t.

	Pain status at time t		
Pain status at time $t-1$	No pain	Moderate pain low interference	Severe pain mod interference
Baseline		One year	
No pain	N/A	N/A	N/A
Moderate/low	0.26	0.66	0.08
Severe/moderate	0.07	0.42	0.51
Severe/high	0.04	0.12	0.83
One year		Two years	
No pain	0.59	0.30	0.11
Moderate/low	0.16	0.73	0.12
Severe/moderate	0.06	0.22	0.72
Two years		Five years	
No pain	0.58	0.32	0.11
Moderate/low	0.24	0.57	0.19
Severe/moderate	0.06	0.28	0.66

The posterior modes of the odds ratios (OR) and the corresponding 95 highest posterior density regions (HPD) from the polytomous regression of the baseline class prevalences on risk factors are shown in Table 3. Depression is strongly associated with baseline pain severity. A unit increase in SCL depression is associated with a 8.7 times higher odds of having *severe pain/high interference* compared to *moderate pain/low interference* (95% HPD = 5.8 to 14.2) and a 3.1 times higher odds of having *severe pain/high interference* compared to *severe pain/moderate interference* (95% HPD = 2.4 to 4.2). In addition, people with lower education and females are more likely to have *severe pain*.

For the regression on the transition probabilities, we were most interested in risk factors for *chronic pain,* which we define as *continued severe pain*. Table 4 shows the corresponding odds ratios of interest, adjusted for age, sex, and education. Since there are two severe pain groups at baseline, separate

Table 3
Relative odds (95% HPD) from regression on baseline class prevalences. Significant values are in bold

Covariate	Severe/high vs. moderate/low		Severe/high vs. severe/moderate		Severe/moderate vs. moderate/low	
Age (+10 years)	1.01	(0.84, 1.22)	1.04	(0.89, 1.22)	0.98	(0.85, 1.12)
Education	**0.52**	**(0.41, 0.66)**	**0.69**	**(0.55, 0.85)**	**0.76**	**(0.65, 0.90)**
Females vs. males	**2.23**	**(1.31, 3.74)**	**1.16**	**(0.75, 1.91)**	**1.84**	**(1.27, 2.66)**
Depression (SCL)	**8.73**	**(5.79, 14.2)**	**3.14**	**(2.36, 4.19)**	**2.93**	**(1.99, 4.17)**

odds ratios are reported for those with *severe pain/moderate interference* and those with *severe pain/high interference* at baseline. Depression level at the previous visit is significantly related to the probability of continued *severe pain* at all years. For example, the odds of remaining in the *severe pain* group compared to transitioning to the *moderate pain* group is 2.8 times greater for a one-unit increase in SCL depression from one to two years (95% HPD = 1.8 to 4.9) and 2.1 times greater from two to five years (95% HPD = 1.3 to 3.6). In addition, the odds of remaining in the *severe pain/moderate interference* group compared to transitioning to the *no pain* group are between 1.75 and 1.99 for all years, though they did not reach statistical significance. Age, sex, and education were not consistent predictors of *chronic pain*.

A limitation of this study is the large amount of missing data in the later years (1 year: 8%, 2 years: 20%, 5 years: 37%). Since our approach is likelihood based, the model assumes the data is missing at random (if the model is correctly specified). This assumption may be plausible for our data, since the study was a very low-impact 30-minute phone interview. In addition, our analysis did not constrain regression coefficients to be constant over time. However, differential dropout by pain severity level could bias the transition probabilities and regression coefficients.

4.4.1 Model diagnostics. We assessed the model assumptions and fit using posterior predictive check distributions (Gelman et al., 1995). The choice of distributions for the observed outcomes was checked by comparing their observed and predicted distributions at each time point using histograms and QQ-plots. The distributions for the three continuous measures are well approximated by the model (Figure 1), except that the left tail of the predicted *pain intensity* distribution is slightly heavier than observed at baseline. This is understandable, since subjects were selected based on seeing their doctor for back pain. The multinomial measure is very well approximated (plots not shown). The number of people *kept from full-time work* across all time points is generally well predicted from the model (Table 5). The posterior predictive p-values (Carlin et al., 2001), i.e., the probability that the predicted values are greater than the observed values, are

Table 4
Relative odds from regression on transition probabilities for patients with severe pain at time $t-1$

	Severe pain vs. no pain		Severe pain vs. moderate pain	
	Mode	(95% HPD)	Mode	(95% HPD)
Baseline to year 1				
Severe pain/moderate interference at baseline:				
Age (+10 years)	1.08	(0.83, 1.50)	1.19	(1.02, 1.40)
Education	0.79	(0.54, 1.10)	**0.69**	**(0.57, 0.83)**
Females vs. males	1.13	(0.55, 2.57)	**1.56**	**(1.02, 2.44)**
Depression (SCL)	1.75	(0.96, 3.51)	1.16	(0.84, 1.62)
Severe pain/high interference at baseline:				
Age (+10 years)	0.79	(0.42, 1.46)	0.84	(0.53, 1.35)
Education	0.85	(0.37, 1.83)	1.68	(0.85, 3.90)
Females vs. males	0.91	(0.13, 5.86)	3.83	(0.82, 19.6)
Depression (SCL)	**3.58**	**(1.23, 11.6)**	1.35	(0.70, 3.06)
Year 1 to year 2				
Age (+10 years)	0.87	(0.64, 1.20)	1.05	(0.85, 1.31)
Education	1.03	(0.69, 1.51)	0.93	(0.71, 1.23)
Females vs. males	1.79	(0.72, 4.37)	1.15	(0.62, 2.15)
Depression (SCL)	1.75	(0.97, 3.29)	**2.79**	**(1.77, 4.89)**
Year 2 to year 5				
Age (+10 years)	1.09	(0.72, 1.67)	1.28	(0.98, 1.71)
Education	0.73	(0.47, 1.18)	**0.69**	**(0.50, 0.93)**
Females vs. males	1.19	(0.36, 3.35)	1.15	(0.52, 2.21)
Depression (SCL)	1.99	(0.92, 4.95)	**2.11**	**(1.32, 3.62)**

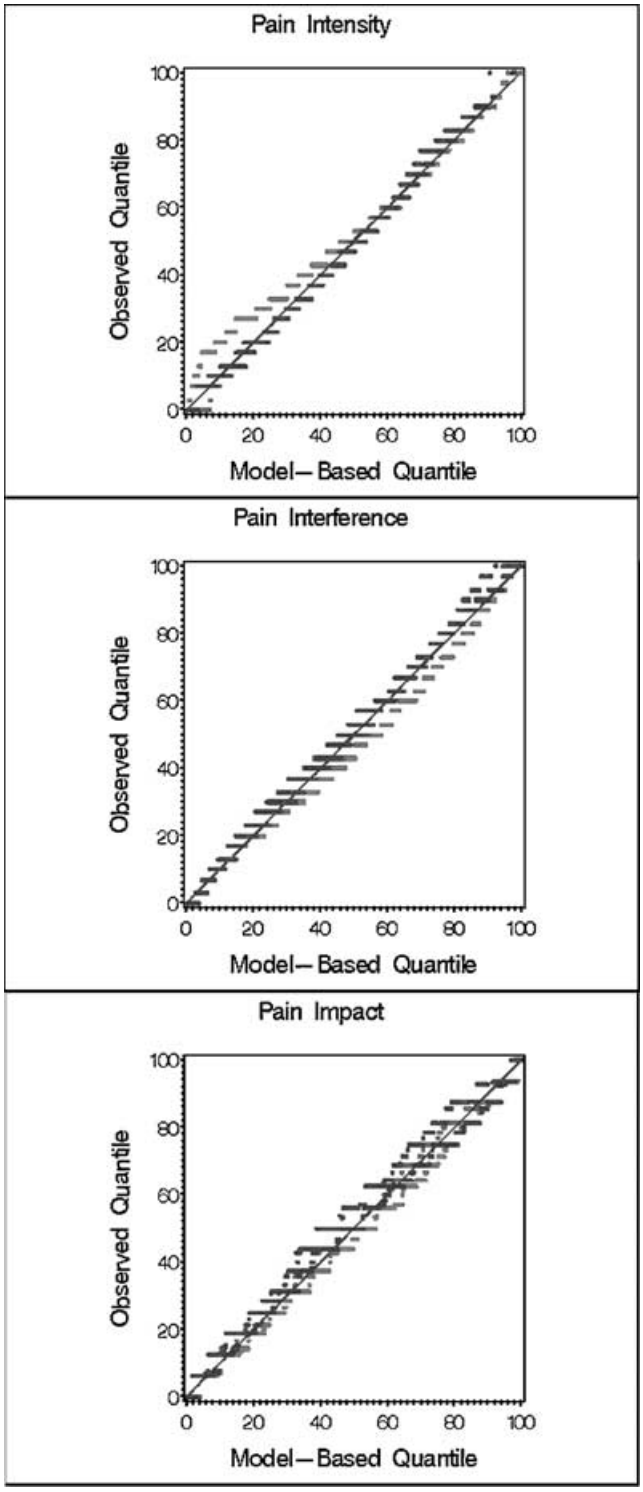

Figure 1. QQ-plots comparing posterior predicted quantiles to observed quantiles for the continuous measures, overlaying time points.

mostly between 0.30 and 0.59; however, the probability of being *kept from full-time work* is slightly overestimated at 1 year ($p = 0.11$) and slightly underestimated at 2 years ($p = 0.96$) among those not retired or homemakers (Table 5).

Table 5
Summary of posterior predictive distribution for the number of people kept from full-time work, including the 2.5, 50, *and* 97.5% *quantiles and the posterior predictive p-value.*

Time point	Observed	2.5%	50%	97.5%	p
Not retired/homemaker					
Baseline	82	64	82	101	0.49
Year 1	90	65	81	97	0.11
Year 2	50	49	63	77	0.96
Year 5	41	29	39	52	0.35
Retired/homemaker					
Baseline	12	6	12	21	0.50
Year 1	22	12	20	29	0.30
Year 2	12	7	13	20	0.59
Year 5	12	7	13	19	0.51

To check the assumption of conditional independence, we plotted the observed versus the predicted Spearman correlations between the five observed outcomes, along with their 95% confidence and HPD intervals (Figure 2). The following symbols indicates the outcome and time point: Ct indicate a correlation between two continuous outcomes at time t, Mt indicates a correlation with the multinomial outcome and a continuous outcome, and Bt indicates a correlation with the binary outcome and either the multinomial outcome or a continuous outcome. Most of the correlations are well approximated by the model, but several are slightly underestimated. These mostly involve correlations with the multinomial outcome.

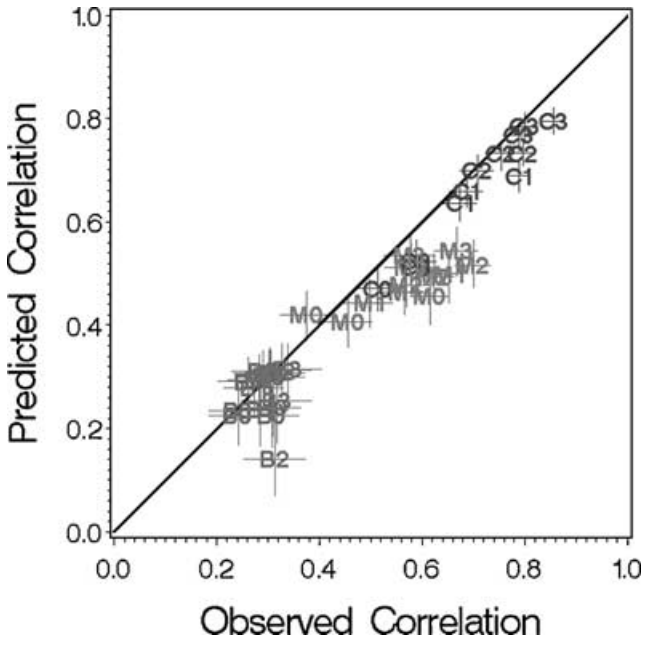

Figure 2. Posterior predicted correlations versus observed correlations, along with 95% confidence and HPD intervals. Symbol indicates outcome type and time point (B = binary, C = continuous, M = multinomial).

5. Discussion

In this article, we introduced a latent transition regression model for longitudinal data with outcomes of mixed types. This approach may be useful when the variable of interest is characterized by multiple continuous, categorical, and count measures, and the interest is in the association among risk factors and the latent pain states over time. We treat health status as a categorical latent variable and model the baseline health status prevalences and transition probabilities between these health states as functions of covariates. We account for residual correlation within the observed measures over time, after conditioning on the latent health status with subject-specific effects. We allow for differential measurement within latent health states by regressing the observed outcomes on covariates in a constrained way. We take a Bayesian model-fitting approach using MCMC.

We chose to model pain status using a categorical, as opposed to a continuous, latent variable, for several reasons. First, it is often natural to summarize health status by observed symptom patterns or subgroups with shared characteristics. Second, when presented with a continuous measure of health status, physicians typically classify individuals into distinct health states using cutoffs, e.g., major disease, mild disease, and no disease). Third, continuous latent variables require additional distributional assumptions that may be more difficult to verify. Finally, we are interested in transitions between different health states over time, and transition models are often easier to interpret with categorical outcomes.

Advantages of using Bayesian sampling methods include flexibility in the choice of distributions (e.g., the truncated normal used in the example) and ease of imposing model constraints, incorporating subject-specific effects, and calculating population average means. A trade-off for this flexibility is the computational intensity of model fitting. Since we use MCMC to fit the models, specialized software is necessary. In addition, latent variable models tend to mix slowly; therefore, tuning parameters for the Metropolis algorithms must be carefully chosen to increase efficiency, and chains typically have high autocorrelation.

One challenge of fitting latent variable models is selecting the numbers of latent health states. One approach is to fit models with different numbers of classes and choose the most parsimonious model that fits the data well using model diagnostics (e.g., Garrett and Zeger, 2000) or some selection criterion such as the Bayesian information criterion (BIC) (Schwarz, 1978). Unfortunately, for models with subject-specific effects, the BIC is not a useful measure (Pauler, 1998). In addition, in the transition setting, models with more than three or four classes become very complicated and are thus less useful scientifically. Therefore, our approach is to fit a model with a scientifically motivated and useful number of classes and then use model diagnostics to guide relaxations in model assumptions. For Bayesian approaches, the posterior predictive check distributions are extremely useful for model checking and revealed that our constrained four-class model generally describes the data quite well. The conditional independence assumption appears to be valid for most of the data, but some of the correlations are underestimated. The potential effect of the residual dependence on inference is not known. Torrance-Rynard and Walter (1997) examined the effect of violations of the conditional independence assumption in the assessment of binary diagnostic test performance. They found that the latent class model often provides unbiased estimates, even when the conditional dependence assumption is violated. They also found that bias is the most severe when class prevalences are small and the conditional outcome prevalences are high. In addition, they found that including one or more conditionally independent outcomes reduces bias. Based on these findings, we do not expect bias in our example, since class prevalences are moderate and the conditional independence assumption appears reasonable for most outcomes. However, more research is needed in this area.

Acknowledgements

This research was supported by National Institute of Health grant P01 DE08773. The author thanks Dr Michael Von Korff for providing the pain data used in the example and for stimulating discussions, and Dr Carolyn Rutter for helpful comments and suggestions. In addition, the author thanks the reviewers and associate editor for their useful comments.

Résumé

L'état sanitaire d'un patient est une observation complexe, souvent caractérisée par plusieurs mesures. L'étude des transitions entre états sanitaires successifs nécessite la prise en compte de multiples variables mesurées longitudinalement. Les difficultés d'analyse de ces données longitudinales multivariées sont encore renforcées lorsque les résultats d'intérêt combinent des variables continues, catégorielles, et des effectifs. Pour résoudre ces difficultés, nous proposons une approche entièrement bayésienne de régression sur les transitions latentes, qui permet l'analyse conjointe d'un mélange de variables longitudinales quelle qu'en soit la distribution. L'état sanitaire est considéré comme une variable catégorielle latente, et les autres variables de résultat sont traitées comme des mesures secondaires de l'état sanitaire latent, avec des erreurs de mesure. Cette approche permet l'estimation conjointe de la prévalence de chaque état sanitaire, et des probabilités de transition entre états, en fonction des covariables. Les résultats observés sont liés aux états sanitaires sous-jacents par des modèles de régression qui incluent: des effets individuels pour tenir compte de la corrélation résiduelle entre mesures répétées au cours du temps, et des covariables pour tenir compte des différentes mesures de l'état sanitaire latent. Nous illustrons la méthode proposée avec des données provenant d'une étude longitudinale sur le mal de dos.

References

Agresti, A. (1984). *Categorical Data Analysis*. New York: John Wiley.

Carlin, J. B., Wolfe, R., Brown, H., and Gelman, A. (2001). A case study on the choice, interpretation, and checking of multilevel models for longitudinal binary outcomes. *Biostatistics* **2,** 397–416.

Derogatis, L. R., Rickels, K., Ulenhuth, E. H., and Covi, L. (1974). The Hopkins symptom checklist: A measure of primary symptom dimensions. In *Psychological Measurements in Psychopharmacology: Problems in Pharmacopsychiatry,* 79–110. Basel: Karger.

Dionne, C., Koepsell, T. D., Von Korff, M., Deyo, R. A., Barlow, W. E., and Checkoway, H. (1997). Predicting long-term functional limitations among back pain patients in primary care settings. *Journal of Clinical Epidemiology* **50,** 31–44.

Dionne, C. E., Von Korff, M., Koepsell, T. D., Deyo, R. A., Barlow, W. E., and Checkoway, H. (1999). A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. *Spine* **24,** 2339–2345.

Garrett, E. S. and Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics* **56,** 1055–1081.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis,* 167–172. London: Chapman and Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.

Gray, S. M. and Brookmeyer, R. (2000). Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete, or time-to-event response variables. *Journal of the American Statistical Association* **95**(450), 396–406.

Huang, G.-H. and Bandeen-Roche, K. (2003). Latent variable regression with covariate effects on underlying and measured variables: An approach of analyzing multiple polytomous surrogates. *Psychometrika,* in press.

Humphreys, K. and Janson, H. (2000). Latent transition analysis with covariates, nonresponse, summary statistics and diagnostics: Modeling children's drawing development. *Multivariate Behavioral Research* **35,** 89–119.

Johnson, L. N. and Kotz, S. (1970). *Continuous univariate distributions,* Volume 1. Distributions in Statistics. New York: Wiley.

Legler, J. M. and Ryan, L. M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association* **92,** 13–20.

Legler, J. M., Lefkopoulou, M., and Ryan, L. M. (1995). Efficiency and power of tests for multiple binary outcomes. *Journal of the American Statistical Association* **90**(430)**,** 680–693.

Lin, H., McCulloch, C., Turnbull, B., Slate, E., and Clark, L. (2000). A latent class mixed model for analyzing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* **19,** 1303–1318.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics* **21,** 1087–1091.

Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55,** 463–469.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40,** 1079–1087.

Pauler, D. K. (1998). The Schwarz criterion and related methods for normal linear models. *Biometrika* **85,** 13–27.

Pincus, T., Burton, A. K., Vogel, S., and Field, A. P. (2002). A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* **27,** 109–120.

Pocock, S. T., Geller, N. L., and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* **43,** 487–498.

Raftery, A. E. and Lewis, S. M. (1996). Implementing MCMC. In *Markov Chain Monte Carlo in Practice,* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), 115–130. London: Chapman and Hall.

Reboussin, B. A., Liang, K.-Y., and Reboussin, D. M. (1999). Estimating equations for a latent transition model with multiple discrete indicators. *Biometrics* **55,** 839–845.

Roeder, K., Lynch, K. G., and Nagin, D. S. (1999). Modeling uncertainty in membership: A case study in criminology. *Journal of the American Statistical Association* **94,** 766–776.

Roy, J. and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56,** 1047–1054.

Rutter, C. R. and Miglioretti, D. L. (2003). Estimating the accuracy of psychological scales using longitudinal data. *Biostatistics* **4**(1), 97–107.

Sammel, M. D. and Landis, J. R. (1998). Summarizing mixed outcomes for pain in interstitial cystitis: A latent variable approach. In *Proceedings of the International Biometric Conference,* pp. 21–30.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52,** 650–663.

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B* **59,** 667–678.

Sammel, M. D., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine* **18,** 2479–2492.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Stastistics* **6,** 461–464.

Torrance-Rynard, V. L. and Walter, S. D. (1997). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine* **16,** 2157–2175.

Von Korff, M., Ormel, J., Keefe, F. J., and Dworkin, S. F. (1992). Grading the severity of chronic pain. *Pain* **50,** 133–149.

Von Korff, M., Deyo, R. A., Cherkin, D., and Barlow, W. (1993). Back pain in primary care: Outcomes at one year. *Spine* **18,** 855–862.

Von Korff, M., Barlow, W., Cherkin, D., and Deyo, R. A. (1994). Effects of practice style in managing back pain. *Annals of Internal Medicine* **121,** 187–195.

Received January 2002. *Revised December* 2002.
Accepted January 2003.

Appendix

We describe the details of the MCMC algorithm used to sample from the joint posterior distribution for the example (Section 4). Each parameter or block of parameters was updated in turn by conditioning on all other parameters in the style of Gibbs sampling. The latent health states, $\mathbf{S}$, regression coefficients, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and subject-specific effect precisions, $\boldsymbol{\zeta}$, were directly sampled from their full conditional distributions as described in Section 3. The remaining measurement model parameters, $\boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \boldsymbol{\delta}$, and $\mathbf{u}$, were updated using random walk Metropolis steps with tuning parameters estimated via the three-simulation strategy of Raftery and Lewis (1996). Thus, we sampled z from a normal distribution with zero mean and standard deviation equal to 2.3 times the estimated full conditional standard deviation and set the candidate value x^* equal to the sum of the current value x and z. The candidate value was accepted with probability equal to the product of the prior distribution and the likelihood given the candidate value, over the product of the prior distribution and the likelihood given the current value.

A.1 *Truncated Normal Outcomes*

For the truncated normal variables, the acceptance probability of the candidate regression coefficient $\boldsymbol{\gamma}^*_{mj}$, $m=1, 2, 3$, is equal to

$$Z_{mj}(\gamma)\exp\left(\sigma^{-2}_{mj}\left[\left(\gamma^*_{mj}-\gamma_{mj}\right)\sum_{i=1}^{N}\sum_{t=0}^{T_i-1}I(S_{it}=j)(y_{imt}-u_{im}) - \frac{N_{jt}}{2}\left\{\left(\gamma^*_{mj}\right)^2-\gamma^2_{mj}\right\}\right]\right),$$

where

$$Z_{mj}(\gamma) = I\left(-10<\boldsymbol{\gamma}^*_{mj}<110\right) \times\prod_{i=1}^{N}\prod_{t=0}^{T_i-1}\left[\frac{\int_0^{100}\sigma^{-2}_{mj}\left\{z-\left(\gamma_{mj}+u_{im}\right)\right\}d\phi(z)}{\int_0^{100}\sigma^{-2}_{mj}\left\{z-\left(\gamma^*_{mj}+u_{im}\right)\right\}d\phi(z)}\right]^{I(S_{it}=j)},$$

$I(\cdot)$ is the indicator function, $\phi(z)$ is the standard normal density, and $N_{jt}=\sum_{i=1}^{N}I(S_{it}=j)$. The acceptance probability for the jth candidate precision value, $(\sigma^{-2}_{mj})^*$, $m=1, 2, 3$, is

$$Z_{mj}(\gamma)\left\{\frac{\left(\sigma^{-2}_{mj}\right)^*}{\sigma^{-2}_{mj}}\right\}^{A+\frac{N_{jt}}{2}-1} \times\exp\left[-\left\{\left(\sigma^{-2}_{mj}\right)^*-\sigma^{-2}_{mj}\right\} \times\left\{B+\frac{1}{2}\sum_{i=1}^{N}\sum_{t=0}^{T_i-1}I(S_{it}=j)(y_{imt}-\mu_{imt})^2\right\}\right],$$

where the hyperparameters $A=B=0.01$. The acceptance probability for the ith candidate subject-specific effect, u^*_{im}, $m=1, 2, 3$, is

$$\exp\left[-\frac{\zeta_m}{2}\left\{\left(u^*_{im}\right)^2-u^2_{im}\right\}\right] \times\prod_{t=1}^{T_i-1}\frac{\int_0^{100}\sigma^{-2}_{mS_{it}}\{z-(\gamma_{mS_{it}}+u_{im})\}\,d\phi(z)}{\int_0^{100}\sigma^{-2}_{mS_{it}}\left\{z-\left(\gamma_{mS_{it}}+u^*_{im}\right)\right\}d\phi(z)} \times\exp\left[\sum_{t=0}^{T_i-1}\sigma^{-2}_{mS_{it}}\left\{\left(u^*_{im}-u_{im}\right)\left(y_{imt}-\gamma_{mS_{it}}\right) -\frac{1}{2}\left(\left(u^*_{im}\right)^2-u^2_{im}\right)\right\}\right].$$

A.2 *Multinomial Outcome*

For the multinomial outcome, the acceptance probability of the jth vector of candidate regression coefficients $\boldsymbol{\gamma}^*_{4j}$ is equal to

$$\exp\left[-2/9\sum_{c=1}^{4}\left\{\left(\gamma^*_{4jc}\right)^2-\gamma^2_{4jc}\right\}\right] \times\prod_{i=1}^{N}\prod_{t=0}^{T_i-1}I\left(S_{it}=j\right)\frac{\left\{\pi\left(\gamma^*_{4j1}\right)\right\}^{I(y_{i4t}=1)}\cdots\left\{\pi\left(\gamma^*_{4j5}\right)\right\}^{I(y_{i4t}=5)}}{\left\{\pi\left(\gamma_{4j1}\right)\right\}^{I(y_{i4t}=1)}\cdots\left\{\pi\left(\gamma_{4j5}\right)\right\}^{I(y_{i4t}=5)}},$$

where $\pi(\gamma_{4jc})=\exp(\gamma_{4jc}+u_{i4c})/\sum_{d=1}^{5}\exp(\gamma_{4jd}+u_{i4d})$ and $\gamma^*_{4j5}=\gamma_{4j5}=u_{i45}=0$. The acceptance probability for the ith vector of candidate subject-specific effects $\mathbf{u}^*_{i4}$ is equal to

$$\exp\left[-\frac{\zeta_4}{2}\sum_{c=1}^{4}\left\{\left(u^*_{i4c}\right)^2-u^2_{i4c}\right\}\right] \times\prod_{t=0}^{T_i-1}\frac{\left\{\pi\left(\gamma^*_{4S_{it}1}\right)\right\}^{I(y_{i4t}=1)}\cdots\left\{\pi\left(\gamma^*_{4S_{it}5}\right)\right\}^{I(y_{i4t}=5)}}{\left\{\pi\left(\gamma_{4S_{it}1}\right)\right\}^{I(y_{i4t}=1)}\cdots\left\{\pi\left(\gamma_{4S_{it}5}\right)\right\}^{I(y_{i4t}=5)}}.$$

A.3 *Binary Outcome*

For the binary outcome, the acceptance probability of the jth candidate regression coefficient, γ^*_{5j}, is equal to

$$\exp\left[-\frac{2}{9}\left\{\left(\gamma^*_{5j}\right)^2-\gamma^2_{5j}\right\}\right] \times\prod_{i=1}^{N}\prod_{t=0}^{T_i-1}I(S_{it}=j)\frac{\left\{\pi\left(\gamma^*_{5j}\right)\right\}^{y_{i5t}}\left\{1-\pi\left(\gamma^*_{5j}\right)\right\}^{1-y_{i5t}}}{\left\{\pi(\gamma_{5j})\right\}^{y_{i5t}}\left\{1-\pi(\gamma_{5j})\right\}^{1-y_{i5t}}},$$

where $\pi(\gamma) = \text{logit}^{-1}\{\gamma + u_{i5} + \delta_1 I(t=0) + \delta_2 z_{it} + \delta_3 z_{it} I(t=0)\}$. Similarly, the acceptance probability for the vector of candidate regression coefficients, $\boldsymbol{\delta}$, is equal to

$$\exp\left[-\frac{2}{9}\sum_{p=1}^{P}\left\{\left(\delta_p^*\right)^2 - \delta_p^2\right\}\right] \times \prod_{j=1}^{4}\prod_{i=1}^{N}\prod_{t=0}^{T_i-1} I(S_{it}=j)\frac{\{\pi(\delta^*)\}^{y_{i5t}}\{1-\pi(\delta^*)\}^{1-y_{i5t}}}{\{\pi(\delta)\}^{y_{i5t}}\{1-\pi(\delta)\}^{1-y_{i5t}}},$$

where $\pi(\delta) = \text{logit}^{-1}\{\gamma_{5j} + u_{i5} + \delta_1 I(t=0) + \delta_2 z_{it} + \delta_3 z_{it} I(t=0)\}$. The acceptance probability for the ith subject-specific effect, u_{i5}^*, is equal to

$$\exp\left[-\frac{\zeta_5}{2}\left\{\left(u_{i5}^*\right)^2 - u_{i5}^2\right\}\right]\prod_{t=0}^{T_i-1}\frac{\left\{\pi\left(u_{i5}^*\right)\right\}^{y_{i5t}}\left\{1-\pi\left(u_{i5}^*\right)\right\}^{1-y_{i5t}}}{\left\{\pi\left(u_{i5}\right)\right\}^{y_{i5t}}\left\{1-\pi\left(u_{i5}\right)\right\}^{1-y_{i5t}}}$$

where $\pi(u_{i5}) = \text{logit}^{-1}\{\gamma_{5S_{it}} + u_{i5} + \delta_1 I(t=0) + \delta_2 z_{it} + \delta_3 z_{it} I(t=0)\}$.