

Longitudinal profiling of health care units based on continuous and discrete patient outcomes

MICHAEL J. DANIELS*

Department of Statistics, University of Florida, Gainesville, FL 32611, USA
mdaniels@stat.ufl.edu

SHARON-LISE T. NORMAND

Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA and
Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA

SUMMARY

Monitoring health care quality involves combining continuous and discrete outcomes measured on subjects across health care units over time. This article describes a Bayesian approach to jointly modeling multilevel multidimensional continuous and discrete outcomes with serial dependence. The overall goal is to characterize trajectories of traits of each unit. Underlying normal regression models for each outcome are used and dependence among different outcomes is induced through latent variables. Serial dependence is accommodated through modeling the pairwise correlations of the latent variables. Methods are illustrated to assess trends in quality of health care units using continuous and discrete outcomes from a sample of adult veterans discharged from 1 of 22 Veterans Integrated Service Networks with a psychiatric diagnosis between 1993 and 1998.

Keywords: Bayesian hierarchical model; Correlation matrix; Informative priors; Latent variable; Mental health.

1. INTRODUCTION

Profiling health care providers involves combining information collected on subjects treated by health care providers and making inferences about the quality of care associated with each provider. The collected information has typically been one of three types: cross-sectional univariate outcomes (Goldstein and Spiegelhalter, 1996; Normand *et al.*, 1997); longitudinal univariate outcomes (Aguilar and West, 1998; Bronskill *et al.*, 2002); or cross-sectional multivariate outcomes (Landrum *et al.*, 2003). In all cases, the outcomes are clustered within health care units. It has become increasingly common to collect both continuous and discrete outcomes longitudinally on providers in order to infer trends in quality of care.

The Veterans Health Administration (VHA) is a case in point. Some have argued that care in the VHA, the largest vertically integrated health care system in the United States, is poorer than in non-VHA institutions (United States Senate, 1999). In 1995, the VHA initiated the National Mental Health Performance Monitoring System and began to publish systematic performance data for the 22 geographically defined service networks that provide mental health services to its veterans. Table 1 summarizes outcomes over a 6-year period for all adult veterans who had an inpatient admission in a general psychiatric program

*To whom correspondence should be addressed.

Table 1. *Characteristics of service network outcomes in the VHA, 1993–1998*

	Fiscal year						All years
	1993	1994	1995	1996	1997	1998	
Total number of patients	38 088	39 574	40 201	38 955	38 627	37 498	23 2943
(<i>k</i>) Mean number of patients	1731	1799	1827	1771	1756	1704	10 589
Inpatient measures							
(1) Mean percent readmitted within 180 days	36.9	35.3	35.2	35.1	34.4	33.5	35.2
Patients with a subsequent admission							
(2) Mean number of readmissions	1.7	1.7	1.7	1.7	1.7	1.7	1.7
(3) Mean number of bed days	33.9	33.3	31.4	29.2	24.4	22.3	29.3
(4) Mean days to first readmission	62.8	64.1	64.0	64.5	63.6	63.4	64.1
Outpatient measures							
(5) Mean percent OP visit within 180 days	74.2	73.7	75.3	77.4	77.6	79.5	76.2
Patients with at least one OP visit							
(6) Mean number of OP visits	12.2	13.3	13.4	14.9	17.6	18.8	14.5
(7) Mean days to first OP visit	35.5	35.4	34.8	33.8	31.8	30.6	33.9
(8) Mean number of periods of continuity	1.6	1.7	1.6	1.7	1.8	2.5	1.8

Numbers represent the mean per service network; OP = outpatient. A fiscal year is defined as October 1 of previous year through September 30 of current year, e.g. fiscal year 1998 corresponds to October 1, 1997 to September 30, 1998. *k* indexes outcome measure.

located in Veterans Affairs Medical Centers during the first 6 months of each fiscal year. Large values of the inpatient measures, with the exception of days to readmission, and small values of the outpatient measures, with the exception of days to first visit, indicate poor quality of care (Rosenheck and DeLilla, 1999). The goal for the VHA is to characterize trends in quality of individual service networks on the basis of the outcomes and to identify problems in quality for specific networks.

The observed data consist of multilevel continuous and discrete outcomes collected repeatedly within clusters over time. Heterogeneity is introduced from several sources: the correlation among the multivariate outcomes on the same patient, correlation among the multivariate outcomes on the same health care unit, the repeated measurement of health units over time, and the heterogeneity across health units. Several methods have been proposed to handle joint modeling of multilevel continuous and discrete outcomes in the cross-sectional setting (Dunson, 2000; Dunson *et al.*, 2003; Gueorguieva and Agresti, 2001; Landrum *et al.*, 2003; Lee and Shi, 2001). The key idea is the introduction of a set of latent variables that reduce the dimensionality of the problem for outcomes having different measurement scales. Fewer methods are available for repeated measurements of multilevel discrete and continuous outcomes. Aguilar and West (1998) proposed a multivariate random-effects time series model for repeated binary outcomes through an auto-regressive structure for the cluster random effects. However, this model is restricted to joint modeling of binary outcomes. Outside the context of multilevel data, methodology for longitudinal data for the latent variable (Dunson, 2003; Roy and Lin, 2000; Oort, 2001), latent class (Lin *et al.*, 2002; Duncan *et al.*, 2002; Muthén *et al.*, 2001; Douglas *et al.*, 1999; Douglas, 1999), and structural equation model (Lee *et al.*, 1992; Mueller, 1996) settings have been developed.

In this article, we consider an analysis of I units measured repeatedly over T time points on the basis of K -dimensional response vectors of continuous and discrete data from a total of N individuals. A model-based approach that distinguishes these features has the advantage of providing estimates that filter the sampling and measurement noise from the underlying quality signals, thus providing more reliable

estimates of quality (McClellan and Staiger, 1999). We propose multilevel multidimensional latent variable models for the multivariate data that permit inclusion of covariates and serial dependence among the latent variables. Underlying normal regression models for each outcome are used and dependence among different outcomes is induced through latent variables. Serial dependence is accommodated through modeling pairwise correlation of the latent variables. We use Bayesian methods for inference. We describe the substantive problem more completely in Section 2 as well as introduce the model and discuss identifiability issues. Section 3 describes our approach to inference, including the prior distributions and posterior computations. In Section 4, we apply our model to the VHA data and make concluding remarks in Section 5.

2. JOINT MODELS FOR LONGITUDINAL MULTILEVEL CONTINUOUS AND DISCRETE DATA

2.1 Profiling VHA mental health treatment quality

The VHA provides health care to more than 4 million veterans annually and is the nation's largest provider of behavioral health services. Like the broader U.S. mental health care system, in the mid-1990s the VHA undertook several initiatives to reduce costs and to increase the number of veterans served (Kizer *et al.*, 2000). In particular, the agency adopted a budget allocation system that gave service networks strong incentives to serve more patients and to reduce per patient costs. Typically, inpatient programs lose resources as a consequence of such health care reorganizations, as there is a shift in the type of treatment provided. A study by Chen *et al.* (2003) found that the shift from inpatient to outpatient mental health care in the VHA over the period 1995–2001 resulted in a 21% decrease in inpatient spending and a 63% increase in outpatient spending.

With such dramatic shifts in care settings, a natural concern relates to whether the quality of mental health services has declined. On the medical side, a study of the impact of VHA reorganization on the quality of care in long-term care units found an increase in pressure ulcer development (Berlowitz *et al.*, 2001). Because regional service networks have strong financial incentives and large administrative control, the need to carefully monitor the quality of care in programs that may be losing resources as a result of the reorganization is especially elevated in a population of patients who may have difficulty navigating the health system.

2.2 Within-network model

Let Y_{itjk} denote the value of the k th response for the j th subject in the i th unit (network) at time t . Subjects are clustered within units, and units are measured repeatedly.

Within-network model for binary outcomes. Assume for binary-valued variables $Z_{itjk} \sim N(\xi_{itjk}, 1)$ such that $Y_{itjk} = I(Z_{itjk} > 0)$, where

$$\xi_{itjk} = \beta_{0tk} + \beta_{1tk}(x_{1itj} - \bar{x}_{1t}) + \cdots + \beta_{Ptk}(x_{Pitj} - \bar{x}_{Pt}) + \phi_{itj} + \eta_{itk}, \quad (2.1)$$

$$\phi_{itj} \sim N(0, \tau_t^2). \quad (2.2)$$

Here, ϕ_{itj} is a subject-specific random effect, η_{itk} , a network random effect, and x_{pitj} ($p = 1, \dots, P$), a subject-specific covariate. Let $k = 1$ and $k = 5$ correspond to the binary random variable for the inpatient and outpatient measures, respectively. The ϕ_{itj} appears in the regression for the two binary responses for each subject inducing a patient-level correlation between the two responses. We specify the distribution of the η_{itk} in Section 2.3.

Within-network model for continuous outcomes. The continuous variables, Y , are grouped into three inpatient ($k = 2, 3, 4$) and three outpatient ($k = 6, 7, 8$) measures, and trivariate normal distributions

are assumed for each. The two sets of continuous variables are conditional on a positive value for the corresponding binary-valued variables, that is, having an inpatient event and an outpatient event. Let $U \in \{\text{in}, \text{out}\}$ with $U = \text{in}$ corresponding to $k = 2, 3, 4$ and $U = \text{out}$ corresponding to $k = 6, 7, 8$. The model for continuous outcomes is

$$Y_{itj}^{(U)} | x_{itj}^{(U)} = 1 \sim N_3(\xi_{itj}^{(U)}, \Sigma_t^{(U)}), \quad (2.3)$$

$$\xi_{itjk} = \beta_{0tk} + \beta_{1tk}(x_{1itj} - \bar{x}_{1t}) + \cdots + \beta_{Ptk}(x_{Pitj} - \bar{x}_{Pt}) + \eta_{itk}, \quad (2.4)$$

where $\Sigma_t^{(U)}$ denotes the covariance of the continuous variables at time t . See Table 1 for an ordered list of the eight patient-level measures.

2.3 Between-network model

The K network effects are assumed to be represented by L latent variables that are a priori independent,

$$\eta_{itk} | \theta_{il}, \lambda_k, \psi_{tk}^2 \sim N(\lambda_{1k}\theta_{i1t} + \cdots + \lambda_{Lk}\theta_{iLt}, \psi_{tk}^2), \quad (2.5)$$

$$\theta_{il} = (\theta_{il1}, \dots, \theta_{ilT})' \sim N_T(0, \mathbf{R}_l), \quad l = 1, 2, \dots, L, \quad (2.6)$$

where $\lambda_k = (\lambda_{1k}, \dots, \lambda_{Lk})'$ is an $L \times 1$ vector of discrimination parameters assumed constant over time. The latent variables, θ_{il} , represent unobserved traits of an individual unit, such as inpatient and outpatient quality, at time t . The variance of the unit random effects, ψ_{tk}^2 , represents heterogeneity among outcomes in a given year not explained by the L latent variables. \mathbf{R}_l denotes a $T \times T$ correlation matrix that permits serial dependence in the L latent variables.

2.4 Serial dependence model for latent variables

We model the within-unit dependency of the outcomes measured at different times through the correlation matrix, \mathbf{R}_l . Our motivation for specifying some structure on \mathbf{R}_l is two-fold: to *predict* latent quality at future times and to increase power through estimation of fewer correlation parameters. The latter consideration is important in our example as there are only 22 networks for estimating a six-dimensional correlation matrix. The complications in developing models for a correlation matrix arise for three reasons: (1) the marginal variances are fixed at one; (2) the matrix needs to be positive definite; and (3) in our setting, the parameters should be interpretable in a longitudinal context.

We consider first-order Markov models. Recall that $\theta_{il} \sim N_T(0, \mathbf{R}_l)$, which implies $\text{Var}(\theta_{ilt}) = 1$, $t = 1, \dots, T$. For a first-order Markov model, we have

$$p(\theta_{ilt} | \theta_{ilj}, j < t) = p(\theta_{ilt} | \theta_{ilt-1}) = N(\rho_{l,t-1}\theta_{ilt-1}, \sigma_{lt}^2).$$

Using iterated expectations, it is easy to show that $\text{Var}(\theta_{ilt}) = 1$ implies that $\rho_{l,t-1}^2 + \sigma_{lt}^2 = 1$ for $l = 1, \dots, L$ and $t = 1, \dots, T$, where $\rho_{l0} = 0$.

Models for $\rho_{l,t-1}$ can be developed subject to the constraint that $\rho_{l,t-1}$ lies in the open interval $(-1, 1)$. In particular, we consider

$$z(\rho_{l,t-1}) = g_l(t-1; \alpha), \quad (2.7)$$

where $z(\cdot)$ is Fisher's z -transformation and $g_l(t-1; \alpha)$ is a smooth function of time parameterized by α for each of the L latent variables; this is a nonstationary, first-order Markov model. The form of these nonstationary models facilitates prediction by providing a mechanism to extrapolate the dependence structure past T . That is, we can use $g_l(\cdot)$ to fill in $\rho_{l,t-1}$ for $t > T$. A simplified stationary version of this model

involves setting $g_l(t-1; \alpha) = z(\rho_l)$ so that $\rho_{l,t-1} = \rho_l$ which implies $\sigma_{l_t}^2 = \sigma_l^2$. The correlation between observations one time unit apart in this model is ρ_l . Markov models of higher order involve severely constrained parameters due to the restrictions on the marginal variance and are difficult to apply in this setting. The models proposed here are similar to the structured antedependence models proposed in Zimmerman and Nunez-Anton (1997) for a covariance matrix.

2.5 Model identifiability

To examine identifiability of parameters of the between-unit model, we approximate the marginal posterior distribution of η_{itk} given only the within-unit model, defined in (2.1–2.4), with

$$\hat{\eta}_i \sim N_{KT}(\eta_i, V_i), \quad i = 1, \dots, I, \quad (2.8)$$

where V_i is assumed known and $\eta_i = (\eta_{i11}, \eta_{i12}, \dots, \eta_{i1T}, \dots, \eta_{iKT})'$. The priors on η_i are given by (2.5)–(2.6), the between-unit model. Integrating over the latent variables, θ_{ilt} , we obtain

$$\eta_i \sim N_{KT}(0, A), \quad i = 1, \dots, I, \quad (2.9)$$

where A is a $KT \times KT$ matrix with K ($T \times T$) blocks on the main diagonal, the k th of which has the form $\sum_{l=1}^L \lambda_{lk}^2 \mathbf{R}_l + \text{diag}(\psi_{k1}^2, \dots, \psi_{kT}^2)$ and with off-diagonal blocks, O_{jk} , of the form, $\sum_l \lambda_{lj} \mathbf{R}_l \lambda_{lk}$.

For each t ,

$$\eta_{i \cdot t} \sim N_K(0, \Lambda \Lambda' + \Psi_{\cdot t}), \quad (2.10)$$

where $\eta_{i \cdot t} = (\eta_{i1t}, \dots, \eta_{iKt})'$, the k th row of $K \times L$ matrix Λ is $(\lambda_{k1}, \dots, \lambda_{kL})'$, and $\Psi_{\cdot t} = \text{diag}(\psi_{1t}, \dots, \psi_{Kt})$. This resembles a standard factor analytic model. Because V_i is known, Λ and $\Psi_{\cdot t}$ will be identified under standard conditions for factor analysis models. We now need to show that \mathbf{R}_l is identified. Consider,

$$\eta_{ik \cdot} \sim N_T\left(0, \sum_{l=1}^L \lambda_{lk}^2 \mathbf{R}_l + \Psi_{k \cdot}\right), \quad (2.11)$$

where $\eta_{ik \cdot} = (\eta_{ik1}, \dots, \eta_{ikT})'$. As the components of $\lambda_{k \cdot}$ and $\Psi_{k \cdot} = \text{diag}(\psi_{k1}, \dots, \psi_{kT})$ are identified from (2.10), \mathbf{R}_l will be identified from the correlation among η_{ik} over time.

2.6 Characterizing networks over time

We consider several quantities for evaluating networks longitudinally using the latent variables. Assuming small values of θ_{ilt} correspond to poorer quality of care, networks worsening over time may be identified through

$$P_{\theta_{il}}^q = P(\theta_{il1} < T_q(\theta_{l1}), \dots, \theta_{ilT} < T_q(\theta_{lT}) | \mathbf{Y}), \quad (2.12)$$

for each i and l . Here, $T_q(\theta_{lt})$ is the q th quantile of the distribution of $\{\theta_{ilt}; i = 1, \dots, I\}$ and \mathbf{Y} denotes the observed data. Similarly, units improving over time may be identified through calculation of

$$P_{\theta_{il}}^q = P(\theta_{il1} > T_q(\theta_{l1}), \dots, \theta_{ilT} > T_q(\theta_{lT}) | \mathbf{Y}). \quad (2.13)$$

The trend may be characterized more formally by estimating and ranking the slope, $B_{\theta_{il}}$. For example, we can compute

$$P(B_{\theta_{il}} < 0 | \mathbf{Y}), \quad (2.14)$$

where $B_{\theta_{il}} = (X'X)^{-1}X'\theta_{il}$, and X is a $2 \times T$ matrix with t th row, $(1, t)$, to provide inference about the rate of change. Alternatively, Spearman's correlation coefficient may be estimated as a nonparametric alternative.

Changes in the relationship between the latent variables may be assessed through the correlation, e.g.

$$\text{Cor}(\theta_{ilt}, \theta_{il't}); \quad t = 1, 2, \dots, T, \quad (2.15)$$

for $l \neq l'$. An overall summary may be represented as $\bar{\theta}_{il} = \sum_l w_l \theta_{ilt}$ for some selected weights, w_l .

3. INFERENCE

3.1 Prior distributions

We adopt a Bayesian approach to inference and specify prior distributions for the hyperparameters. We assume the presence of two latent variables, one representing the quality of inpatient care and one the quality of outpatient care. This assumption was based on a variety of considerations. First, several studies have shown that the type of health care reorganization operating at the time of our study has generally resulted in different quality changes in the inpatient and outpatient settings. This shifting of inpatient to outpatient treatment settings has been observed in managed care organizations (Mark and Coffey, 2000). Second, subjective measures of quality of mental health services in the VHA, such as patient satisfaction, have demonstrated differences in inpatient and outpatient quality for some subgroups (Hoff *et al.*, 1998). Third, in a prior cross-sectional analysis of the 1998 data, we found empirical evidence of two latent variables (Landrum *et al.*, 2003).

We place informative priors on the discrimination parameters $\boldsymbol{\lambda} \sim N(\boldsymbol{\pi}, c^2 I_{LK})$, where

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}_{1k} \\ \boldsymbol{\pi}_{2k} \end{pmatrix}' = \pi^* \begin{pmatrix} -1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 & -1 \end{pmatrix}'. \quad (3.1)$$

We discuss the choices of $\pi^* > 0$ and c^2 in Section 4.2 and also examine sensitivity to these choices. This is the prior for the discrimination parameters in a two-latent variable model ($L = 2$) and is consistent with an inpatient and an outpatient quality variable. The 1's and -1's in the prior specification correspond to that response's (of the K) contribution to the inpatient and outpatient latent variable. For example, favorable inpatient characteristics include many days to first readmission but few number of bed days; in (3.1), these get the multipliers 1 and -1, respectively. The components of the prior means for $\boldsymbol{\lambda}$ were selected in order to place equal weights on the inpatient and outpatient components as this was deemed most reasonable by the investigator who collected the data. Informative priors on the discrimination parameters or restrictions on the form of the discrimination matrix, Λ , are necessary for identifiability of the latent variables (Lopes and West, 2004).

Diffuse proper priors are specified for the regression coefficients, β_{ik} , i.e. normal priors with mean 0 and variance 10 000 and gamma priors on the inverse of the variance components, $1/\tau_i^2$ and $1/\psi_{ik}^2$, with parameters (0.001, 0.001). A Wishart prior is placed on $\Sigma_i^{(U)-1}$ with degrees of freedom equal to $\dim(\Sigma_i^{(U)})$ and scale matrix equal to the identity matrix. Because the continuous responses were standardized, marginal variances of one for the scale matrix in the Wishart prior seemed reasonable. For the unstructured correlation matrix, \mathbf{R}_l , similar to Barnard *et al.* (2000) and Daniels (2005), we specify a uniform prior over the compact subspace of the $T(T-1)/2$ dimensional cubic, $[-1, 1]^{T(T-1)/2}$, such that \mathbf{R}_l is positive definite. A uniform prior on the interval $(-1, 1)$ is specified for the correlation parameter ρ_l in the Markov model given in Section 2.3.

3.2 Sampling and model fit

A Gibbs sampler was designed to sample from the posterior distribution of the parameters. The full conditional distributions are all known forms (gamma distributions for the inverse of the variance components, normal distributions for the regression parameters, latent variables, and discrimination parameters, and

Wishart distributions for the covariance matrices) except for the correlation matrices, \mathbf{R}_l , $l = 1, \dots, L$. We implement an adaptation of the component-wise approach of Barnard *et al.* (2000) to sample \mathbf{R}_l when \mathbf{R}_l is assumed to be unstructured. In the case of Markov dependence, we use a random walk Metropolis–Hastings algorithm to sample the correlation parameter.

We sampled the parameters in blocks to increase the efficiency of the sampling algorithm. In particular, we sampled the following parameters as blocks in the within-network model: the $P \times 1$ vectors β_{tk} , for $k = 1$ and $k = 5$; the $(3P) \times 1$ vectors of $(\beta_{t2}, \beta_{t3}, \beta_{t4})$ and $(\beta_{t6}, \beta_{t7}, \beta_{t8})$; the matrices $\Sigma_t^{(U)}$ for U and t ; and in the between-network model, the $L \times 1$ vector of λ_k for $k = 1, \dots, K$, and the $T \times 1$ vectors θ_{il} for $l = 1, \dots, L$. The other parameters were sampled component-wise.

Model comparison and fit. We use the deviance information criterion (DIC) of Spiegelhalter *et al.* (2001) as an overall measure of model fit. Because the main unit of interest is the network, we utilize the following integrated likelihood as the basis for the deviance in the DIC

$$\prod_t \iiint \prod_i \prod_j p(y_{itj} | \eta_{it}, \beta_t, \phi_{itj}, \Sigma_t) p(\phi_{itj} | \tau_t^2) p(\Sigma_t) p(\beta_t) p(\tau_t^2) d\phi_{itj} d\beta_t d\tau_t^2 d\Sigma_t,$$

where y_{itj} is the $K \times 1$ dimensional response vector and η_{it} is a $K \times 1$ dimensional latent variable representing K outcomes at the network level for each time, t . Because this is not available in closed form and the number of patients per network is quite large, we approximate the integrated likelihood using

$$\hat{\eta}_i \sim N_{KT}(\eta_i, V_i), \quad i = 1, \dots, I, \quad (3.2)$$

where $\hat{\eta}_i$ is the maximum likelihood estimate (MLE) of $\eta_i = \{\eta_{itk}\}$ (here, we use posterior mean) and V_i is the estimated variance (posterior variance). We computed this approximation by fitting the within-unit model, given by (2.1)–(2.4) using diffuse normal priors on the η_{itk} , i.e. normal priors with mean 0 and variance 10 000. Specifically, we do not use the between-unit model, (2.5)–(2.6), when constructing this approximation to the network likelihood. We also note that V_i is a block diagonal because of the assumed independence across years, and our data in this setting are the $\hat{\eta}_i$. Results in Daniels and Kass (1998) can be used to show that the approximation in (3.2) has the same order of error as a Laplace approximation, here $O(\sum_i \sum_t 1/n_{it})$, where n_{it} is the number of patients in network i in year t , for estimating the between-network parameters; in addition, results given in that paper show that the order of the error is unchanged by replacing the MLE with the posterior mean. Finally, the network-level data have been adjusted for patient mix and the correlation among the multivariate outcomes for each patient.

We use the DIC to aid in the choice of an appropriate form for the correlation matrices although it could also be used to choose the number of latent variables in the between-unit model (2.5)–(2.6).

We also use this approximation to conduct posterior predictive checks (Gelman *et al.*, 2003) for assessing model fit. In this context, we determine whether the features of the posterior distribution of η_{itk} from fitting the full model are consistent with the data, $\hat{\eta}_{itk}$, which are based on (3.2). Our goal is to assess the fit of the *between-network model*. Because the network-level ‘data’ constructed using the approximation given by (3.2) did not use the between-network model, we will not obtain overly optimistic results from the posterior predictive checks. This approach is somewhat nonstandard as the $\hat{\eta}$ are not observed data—the y_{itj} are observed. To do this, we sample a set of $\hat{\eta}'$ using (3.2) given the current values of the parameters, $\eta_i^{(r)}$, sampled from $\eta_i | y$ using the Gibbs sampler.

We examine two correlations: (1) the correlation of the components of η_i across k given t , $\text{corr}(\eta_{ikt}, \eta_{ik't})$ and (2) the correlation of the components of η_i across t given k , $\text{corr}(\eta_{ikt}, \eta_{ikt'})$. For each, we compute a single summary based on the data, $\hat{\eta}_{itk}$, and a distribution of the quantity from the posterior predictive distribution using $\hat{\eta}'_{itk}$. We summarize these checks via posterior predictive p -values. If the

model fits well, we expect the p -values to not be extreme (e.g. close to 0 or 1) as this would indicate that the particular feature of the data is not captured well by the model.

At each iteration of the Gibbs sampler, we also calculate a confidence region given by $(\hat{\eta}_i - \eta_i^{(r)})' V_i^{-1} (\hat{\eta}_i - \eta_i^{(r)})$. A 95% cutoff for this region is calculated based on a χ^2 distribution which should be approximately correct at each iteration of the sampler, given the parameters values at that iteration. We evaluate an empirical coverage probability by averaging across iterations. Because $\hat{\eta}_i$ is high dimensional (48), we examine the coverage for both this 48-dimensional vector and also for the 8-dimensional vector of $\eta_{i,t}$ at each time, $t = 1, \dots, T$.

4. APPLICATION TO THE VHA STUDY

4.1 Details on the VHA National Mental Health Performance Monitoring System data

The continuous inpatient and outpatient responses were standardized to have mean 0 and variance 1. Patient-level covariates included in the within-unit model [see (2.1) and (2.4)] included age (categorized as ≤ 39 , 40–59, or ≥ 60), race (categorized as white, black, or other), and primary mental health diagnosis (schizophrenia, other psychoses, posttraumatic stress disorder, alcohol abuse, or other). Over the 6-year period, the majority (95%) of patients were male; one-quarter black; 60% aged between 40 and 59 years of age at the time of their index discharge; and 27% had a primary mental health diagnosis of schizophrenia (data not shown). Table 1 indicates a 3.4% absolute decline between 1993 and 1998 in 6-month readmission rates, and a corresponding 5.3% increase in the outpatient visit rate over the 22 networks. These changes are consistent with expected patterns following health care reorganization. In what follows, networks are labeled using roman letters (A–W).

4.2 Sampling algorithm

We ran five parallel chains, each of length 1100. The chains appeared to have converged within 100 iterations when examining the trace plots of selected parameters within each chain; we therefore used the 100 iterations as the burn-in. To minimize autocorrelation within the chains, we sampled every 10th iteration. The total posterior sample size was 500.

4.3 Prior for discrimination matrix, Λ

We fit a two-latent variable model ($L = 2$) where, a priori, the latent variables were thought to represent inpatient and outpatient quality. The prior specification given in Section 3.1 required specification of two hyperparameters, π^* and c^2 . We wanted to specify priors flexible enough to verify the a priori belief of an inpatient and outpatient latent variable, but informative enough to avoid identifiability problems. We first set $\pi^* = 0.25$ and $c^2 = 0.01$ ($c = 0.1$) and then examined sensitivity to fixing π^*/c and varying π^* ; inferences were basically insensitive to this variation. We then fixed π^* and varied the ratio. Values of π^*/c less than two were getting close to identifiability problems with components of the latent variables switching and signs changing on the discrimination parameters. Values of the ratio over three seemed to be too informative. Thus, for inferences in Sections 4.4–4.6, we set $\pi^* = 0.25$ and $\pi^*/c = 2.5$.

The posterior means of the discrimination parameters were consistent with the prior which differentiated inpatient and outpatient measures (Table 2), although they were considerably smaller than their prior means. The discriminating abilities of the components of the inpatient and outpatient latent variables were not very consistent with equality among the four inpatient and outpatient variables. For example, the discrimination parameters for the number of visits and periods of continuity were larger than visits within 180 days and days to first visit. In addition, the posterior variances of the discrimination parameters were much tighter than the prior variances, indicating that the data very much determined the posterior on λ .

Table 2. *Posterior means and 95% credible intervals for the discrimination parameters, λ_k for the inpatient and outpatient latent variables for the final model given in Section 4.3*

Outcome	Inpatient latent variable	Outpatient latent variable
Readmission within 180 days [†]	−0.06 (−0.09, 0.03)	−0.00 (−0.05, 0.05)
Number of readmissions [†]	−0.03 (−0.08, 0.05)	−0.01 (−0.09, 0.06)
Number of bed days [†]	−0.08 (−0.11, −0.04)	−0.00 (−0.04, 0.04)
Days to first readmission	0.04 (0.02, 0.06)	−0.02 (−0.08, 0.04)
Visit within 180 days [†]	−0.01 (−0.05, 0.03)	−0.07 (−0.10, −0.04)
Number of visits [†]	0.02 (−0.01, 0.06)	−0.12 (−0.15, −0.08)
Days to first visit	0.00 (−0.05, 0.05)	0.06 (0.04, 0.08)
Periods of continuity [†]	−0.01 (−0.04, 0.02)	−0.10 (−0.13, −0.06)

[†]Smaller values correspond to better quality of care.

Table 3. *Posterior mean of unstructured correlation matrices for inpatient (\mathbf{R}_1) and outpatient (\mathbf{R}_2) latent variables*

Year	Latent					
	1993	1994	1995	1996	1997	1998
Latent variable 1: Inpatient quality						
1993	1.000	0.61	0.42	0.30	0.23	0.23
1994		1.000	0.55	0.43	0.35	0.32
1995			1.000	0.37	0.36	0.44
1996				1.000	0.51	0.43
1997					1.00	0.56
Latent variable 2: Outpatient quality						
1993	1.000	0.74	0.75	0.66	0.48	0.21
1994		1.000	0.73	0.59	0.54	0.36
1995			1.000	0.70	0.60	0.34
1996				1.000	0.68	0.41
1997					1.000	0.62

Finally, there was no evidence that the discrimination parameters varied over time implying no need to index λ_k by t .

4.4 Serial dependence

To assess the importance of modeling the dependence of the networks over time, we first fit unstructured correlation matrices for both latent variables, \mathbf{R}_l , $l = 1, 2$. The posterior mean of the correlation matrices for the inpatient and outpatient latent variables in the unstructured model are given in Table 3.

In general, the posterior means of the lag 1 correlations were around 0.5–0.7. The correlation generally decreased as the lag increased. The posterior distribution of the correlations were quite skewed with the posterior medians and mode for the large positive correlations closer to one than the posterior means (Figure 1). A first-order Markov model seemed most reasonable for \mathbf{R}_1 . Because the lag k correlations did not vary much over time (moving down each off-diagonal of the correlation matrix), we fit a first-order

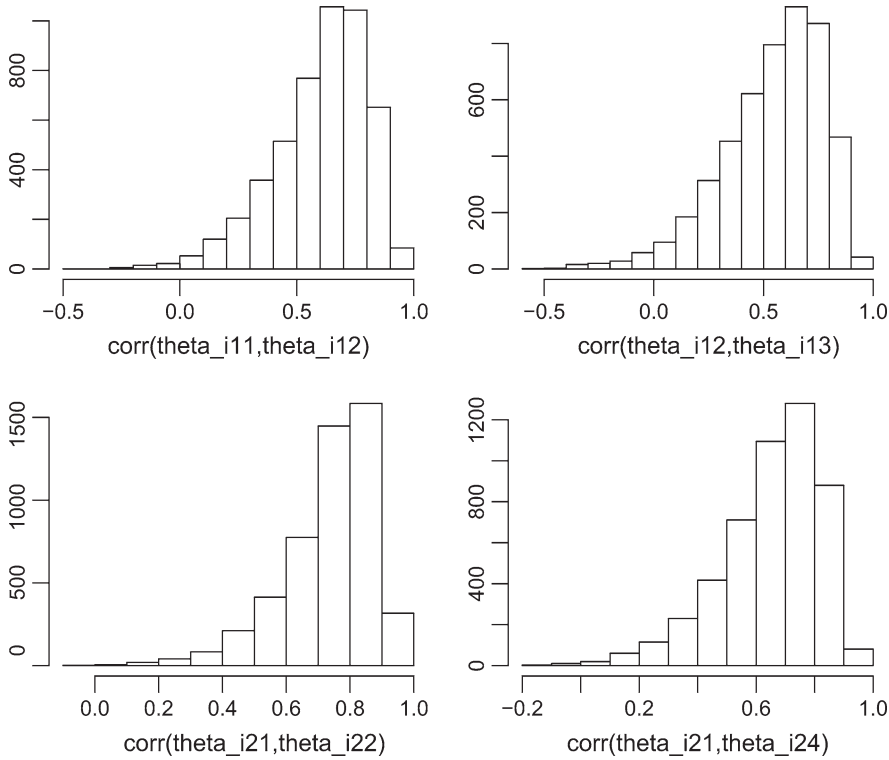


Fig. 1. Posterior distributions of several of the components of the \mathbf{R}_l matrices under the unstructured \mathbf{R} models.

Table 4. *DIC. Markov(1) corresponds to the first-order Markov model discussed in Section 2.4*

Correlation model	DIC	p_D
Independence, $\mathbf{R} = I$	2042	842
Unstructured, \mathbf{R}	1941	777
Markov(1), \mathbf{R}_m	1935	763
Markov(1), Unstructured, $\mathbf{R}_1 = \mathbf{R}_m, \mathbf{R}_2 = \mathbf{R}$	1924	769

Markov model with $\rho_{l,t-1} = \rho_l$. In this first-order model, the lag 1 correlations was quite high with a posterior mean (95% credible interval) of 0.71 (0.63, 0.79).

Ultimately, in addition to the unstructured and Markov specification of the correlation matrix, we also considered independence over time, $\mathbf{R}_l = I$. The $\mathbf{R} = I$ model was fit to assess the potential efficiency gains over modeling the data separately by year as this model did not seem feasible when examining the estimated correlation matrices in Table 3.

Using the DIC, models that employed serial dependence for the inpatient latent variable fit best (Table 4). Fewer effective number of parameters in the dependence models which have more parameters in the \mathbf{R}_l matrices is a result of more shrinkage of the η_i in these models. We used the best-fitting model, $\mathbf{R}_1 = \mathbf{R}_m, \mathbf{R}_2 = \mathbf{R}$, for the basis of inference. An alternative to picking the ‘best’ model would involve model averaging over the various specifications of the correlation matrix. We discuss this in Section 5.

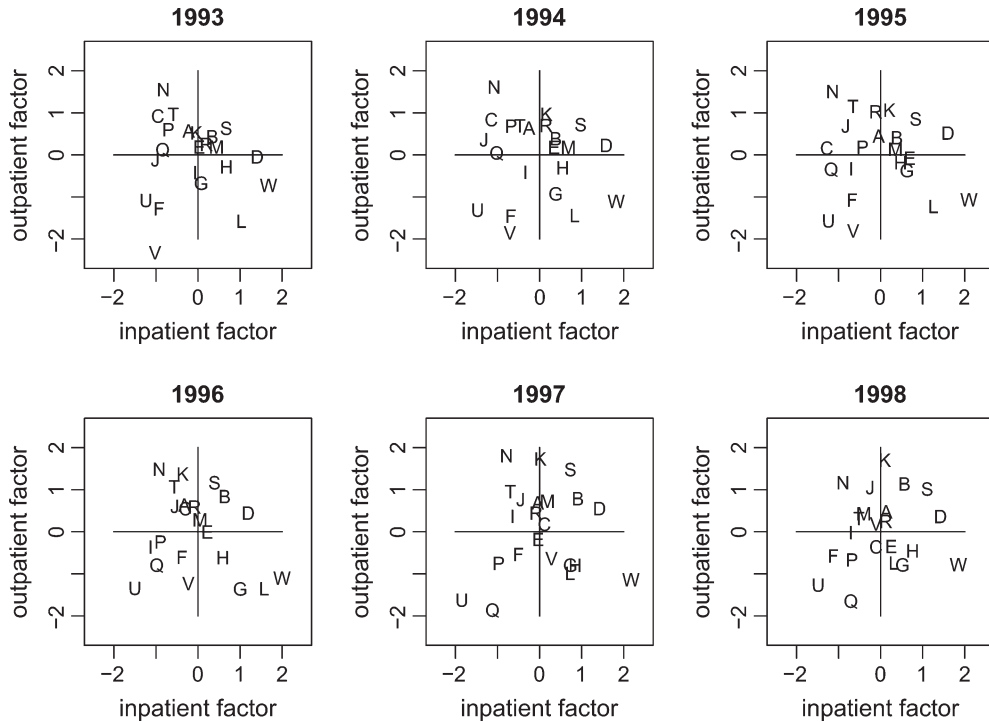


Fig. 2. Posterior means for inpatient and outpatient latent variables for each year.

4.5 Longitudinal patterns of quality

Figure 2 displays scatterplots of the posterior means of the inpatient and outpatient latent variables for each network. These variables summarize the inpatient and outpatient quality of the networks over time with higher values corresponding to better care. Network W had consistently high inpatient quality, network N had consistently high outpatient quality, and network U had consistently poor inpatient and outpatient quality (lower left quadrant) across all years.

Figure 3 displays the posterior means of selected inpatient and outpatient latent variables for a subset of the 22 networks. Some networks appear to be ‘significantly’ improving or worsening over the 6-year period. For example, network V appears to provide improving outpatient quality while network P appears to have worsening outpatient quality.

As a comparison, the probabilities in parentheses below are derived from assuming independence over time, the $\mathbf{R} = I$ model. There was no strong evidence that any of the networks provided consistently poor care over time using (2.12). Setting $q = 0.20$ (20th percentile), the highest probability of being in the lower 20% for all 6 years was network U on inpatient quality with probability 0.49 (0.20 under $\mathbf{R} = I$) and networks U and L on outpatient quality, with probabilities 0.37 (0.04 under $\mathbf{R} = I$) and 0.18 (0.04 under $\mathbf{R} = I$), respectively. In terms of excellent inpatient quality (2.13), network W was in the upper 20% for all 6 years for the inpatient latent variable with probability 0.89 (0.54 under $\mathbf{R} = I$) while network D had probability 0.51 (0.15 under $\mathbf{R} = I$). For excellent outpatient quality, network N had probability 0.52 (0.14 under $\mathbf{R} = I$). From these results, it is clear that inferences change considerably under $\mathbf{R} = I$ model and stronger inferences are possible about the behavior of the networks over time by accounting for the temporal correlation.

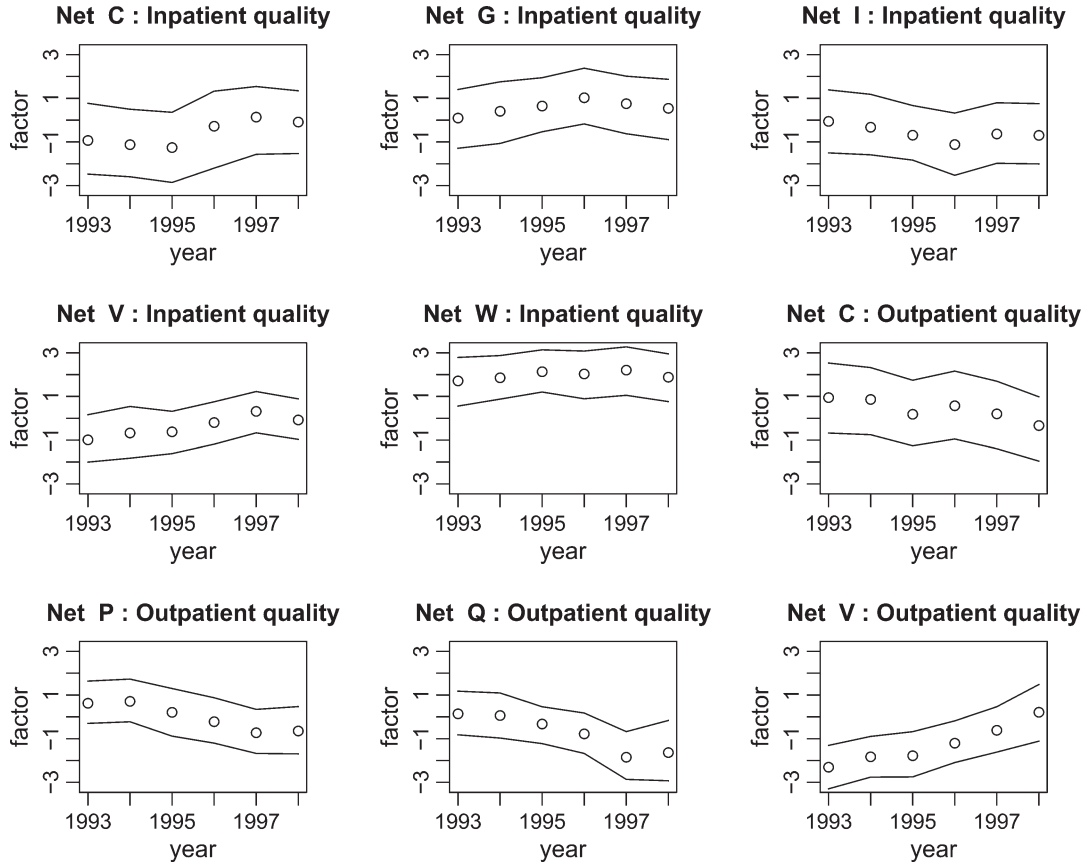


Fig. 3. Posterior means and 95% pointwise credible intervals for selected inpatient and outpatient latent variables over time.

As a less stringent criterion for longitudinal performance, we computed posterior probabilities of a positive (negative) slope for each latent variable for each network as given by (2.14). Networks C, J, and V had posterior probabilities of 0.85 or greater for improving inpatient quality over the 6-year period while network M had a similar probability associated with worsening inpatient quality. Using the same criterion, networks F, K, L, and V had large probabilities for outpatient quality improvement. Degrading outpatient quality trends were associated with networks C, O, and P.

4.6 Posterior predictive checks

We first examined the correlation among the η_i corresponding to each of the eight responses, for each time, t , $\text{corr}(\eta_{ikt}, \eta_{ikt'})$, in order to ensure the model characterized the correlation among the network effects. This check resulted in 168 p -values with only 3 outside the interval (0.05, 0.95) and none outside the interval (0.01, 0.99). We then examined the correlation among the η_i over time for each of the eight responses, $\text{corr}(\eta_{ikt}, \eta_{ikt'})$, to assess the appropriateness of temporal correlations. This check resulted in 120 p -values. The fit was not as good as the previous check, but not unreasonable—24 of the 120 p -values were outside the interval (0.05, 0.95), but only 5 were outside (0.01, 0.99). There was no consistent pattern in terms of particular correlations not being captured well. In addition, for the extreme p -values, the

observed correlation was typically quite close to the observed extreme of the posterior predictive distribution. Finally, the average coverage for the confidence regions described in Section 3.2 for the entire η_i vector was 0.86 and for the lower dimensional $\eta_{i:k}$ vectors, 0.92, which is quite good given the complexity of the data and the ‘parsimonious’ hierarchical model.

5. CONCLUDING REMARKS

We have proposed a multilevel multidimensional model to characterize units measured repeatedly on the basis of within-unit continuous and discrete outcomes. The aggregation level of interest, the health care network, was fit fairly well by the model as assessed by posterior predictive checks. We also observed an increase in efficiency in terms of longitudinal profiling by exploiting the temporal correlation in our model for the data. In our problem, subjects were not measured repeatedly over time, rather the health care unit was measured repeatedly. We introduced temporal correlation at the network level through modeling the pairwise correlations of the latent variables. An alternative approach could involve reparameterization of the matrix from pairwise correlations to partial correlations (Wong *et al.*, 2003). We note however that the correlation structure of longitudinal data is difficult to model parsimoniously using partial correlations and additional constraints would be required.

An important area where more research is needed relates to model fit. Given the complexity of the model, we used an approximate network-level likelihood to evaluate the DIC. Moreover, we focused on the fit of the model at the network level. In complex models, such as those we proposed here, there are several steps where choices are made: the number of latent variables underlying the observed data, the structure of the serial correlation, and the inclusion of covariates. While we did not include unit (network)-level covariates, these could be easily accommodated in the dependence model for the latent variables.

In this paper, we fixed the number of latent variables based on prior research, theory, and expert opinion. An extension would involve accounting for uncertainty in the number of latent variables through a model-averaging approach. However, given the goals of this paper, such an approach would greatly hinder interpretation. As pointed out by a referee, an interesting extension would involve model averaging over the different covariance structures for the latent variables. To accomplish this, methods to compute the marginal distribution of the data would need to be developed and ideally a broader class of covariance models considered.

ACKNOWLEDGMENTS

This research was supported by Grants CA85295 (Daniels) from the National Cancer Institute and MH61434 (Normand) from the National Institute of Mental Health. The VHA data were generously provided through the efforts of Robert Rosenheck, MD, of West Haven, CT. The authors are grateful to Myung Joon Kim for programming expertise.

REFERENCES

- AGUILAR, O. AND WEST, M. (1998). Analysis of hospital quality monitors using hierarchical time series models. In Gatsonis, C., Kass, R. E. *et al.* (eds), *Case Studies in Bayesian Statistics*, Volume 4. New York, NY: Springer, pp. 287–302.
- BARNARD, J., MCCULLOCH, R. AND MENG, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10**, 1281–1311.
- BERLOWITZ, D. R., YOUNG, G. J., BRANDEIS, G. H., BORIS, K. AND ANDERSON, J. J. (2001). Health care reorganization and quality of care: unintended effects on pressure ulcer prevention. *Medical Care* **39**, 138–146.
- BRONSKILL, S., NORMAND, S.-L., LANDRUM, M. B. AND ROSENHECK, R. (2002). Longitudinal profiles of health care providers. *Statistics in Medicine* **21**, 1067–1088.

- CHEN, S., SMITH, M. W., WAGNER, T. D. AND BARNETT, P. G. (2003). Spending for specialized mental health treatment in the VA: 1995–2000. *Health Affairs* **22**, 256–263.
- DANIELS, M. J. (2005). Bayesian modelling of several covariance matrices and some results on the propriety of the posterior for linear regression with correlated and/or heterogeneous errors. *Journal of Multivariate Analysis* (in press).
- DANIELS, M. J. AND KASS, R. E. (1998). A note on first stage approximation in two stage hierarchical models. *Sankhya, Series B*. **60**, 19–30.
- DOUGLAS, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. *Statistics in Medicine* **18**, 2917–2931.
- DOUGLAS, J. A., KOSOROK, M. R. AND CHEWNING, B. A. (1999). A latent variable model of discrete multivariate psychometric waiting times. *Psychometrika* **64**, 69–82.
- DUNCAN, T. E., DUNCAN, S. C., OKUT, H., STRYCKER, L. A. AND LI, F. (2002). An extension of the general latent variable growth modeling framework to four levels of the hierarchy. *Structural Equation Modeling* **9** 303–326.
- DUNSON, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, B* **62**, 355–366.
- DUNSON, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* **98**, 555–563.
- DUNSON, D. B., CHEN, Z. AND HARRY, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59**, 521–530.
- GELMAN, A., CARLIN, J. B., STERN, H. S. AND RUBIN, D. B. (2003). *Bayesian Data Analysis*, 2nd edition. Chapman and Hall.
- GOLDSTEIN, H. AND SPIEGELHALTER, D. J. (1996). Statistical aspects of institutional performance: league tables and their limitations (with discussion). *Journal of the Royal Statistical Society, A* **159**, 385–444.
- GUEORGUEVA, R. V. AND AGRESTI, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102–1112.
- HOFF, R. A., ROSENHECK, R. A., MATERKO, M. AND WILSON, N. J. (1998). The quality of VA mental health services. *Administration and Policy in Mental Health* **26**, 45–56.
- KIZER, K., DEMAIS, J. AND FEUSSNER, J. (2000). Reinventing VA health care: systematizing quality improvement and quality innovation. *Medical Care* **38** (Suppl 1), I7–I16.
- LANDRUM, M. B., NORMAND, S. L. T. AND ROSENHECK, R. A. (2003). Selection of related multivariate means: monitoring psychiatric care in the Department of Veterans Affairs. *Journal of the American Statistical Association* **98**, 7–16.
- LEE, S. Y., POON, W. Y. AND BENTLER, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika* **57**, 89–105.
- LEE, S. Y. AND SHI, J. Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics* **57**, 767–794.
- LIN, H., TURNBULL, B., MCCULLOCH, C. AND SLATE, E. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* **97**, 53–65.
- LOPES, H. F. AND WEST, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67.
- MARK, T. AND COFFEY, R. (2000). Spending on mental health and substance abuse treatment, 1987–1997. *Health Affairs* **July/August**, 108–120.

- MCCLELLAN, M. AND STAIGER, D. (1999). The quality of health care providers. *NBER Working Paper 7327*. Cambridge, MA: National Bureau of Economic Research.
- MUELLER, R. O. (1996). *Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS*. New York: Springer.
- MUTHÉN, B., BROWN, C. H., MASYN, K., JO, B., KHOO, S.-T., YANG, C.-C., WANG, C.-P., KELLAM, S. G., CARLIN, J. B. AND LIAO, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–475.
- NORMAND, S.-L., GLICKMAN, M. E. AND GATSONIS, C. A. (1997). Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association* **92**, 803–814.
- OORT, F. J. (2001). Three-mode models for multivariate longitudinal data. *The British Journal of Mathematical and Statistical Psychology* **54**, 49–78.
- ROSENHECK, R. AND DELILLA, D. (1999). *Department of Veterans Affairs National Mental Health Program Performance Monitoring System: Fiscal Year 1998 Report*. Northeast Program Evaluation Center, VA Health Services Research & Development Service. West Haven, CT: VA Connecticut Healthcare System.
- ROY, J. AND LIN, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56**, 1047–1054.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–616.
- UNITED STATES SENATE (1999). Hearing of ‘quality of care in the Veterans Affairs health care system’, September 22, 1998. *One Hundred Fifth Congress, Second Session*. Washington, DC: Government Printing Office, p. 62.
- WONG, F., CARTER, C. K. AND KOHN, R. (2003). Efficient estimation of covariance selection models. *Biometrika* **90**, 809–830.
- ZIMMERMAN, D. AND NUNEZ-ANTON, V. (1997). Structured antedependence models for longitudinal data. In Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G. and Wolfinger, R. D. (eds), *Modelling Longitudinal Data and Spatially Correlated Data: Methods, Applications and Future Directions*. Springer, pp. 63–76.

[Received August 16, 2004; revised May 6, 2005; accepted for publication May 20, 2005]