Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction

MICHAEL R. ELLIOTT*
Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 612 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA
melliott@cceb.upenn.edu

JOSEPH J. GALLO
Department of Family Practice and Community Medicine, University of Pennsylvania School of Medicine, USA

THOMAS R. TEN HAVE
Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, USA

HILLARY R. BOGNER
Department of Family Practice and Community Medicine, University of Pennsylvania School of Medicine, USA

IRA R. KATZ
Department of Psychiatry, University of Pennsylvania School of Medicine, USA

SUMMARY

Positive and negative affect data are often collected over time in psychiatric care settings, yet no generally accepted means are available to relate these data to useful diagnoses or treatments. Latent class analysis attempts data reduction by classifying subjects into one of \( K \) unobserved classes based on observed data. Latent class models have recently been extended to accommodate longitudinally observed data. We extend these approaches in a Bayesian framework to accommodate trajectories of both continuous and discrete data. We consider whether latent class models might be used to distinguish patients on the basis of trajectories of observed affect scores, reported events, and presence or absence of clinical depression.

Keywords: Cardiovascular disease; Depression; DIC; General growth mixture modeling; Gibbs sampling; Label switching; Model choice.

*To whom correspondence should be addressed.
1. INTRODUCTION

Psychiatric illnesses are characterized by symptoms or behaviors that persist or recur over time. For example, the cardinal features of depressive disorders include the presence of negative and absence of positive moods for periods of several weeks or more (American Psychiatric Association, 2000). Symptoms are usually evaluated by obtaining retrospective histories from patients or other informants in clinical or research assessments. An alternative approach increasingly used in research is to ask subjects to rate their affect (measures of positive or ‘happy’ mood and negative or ‘sad’ mood) through time using daily diaries rather than depend on recall of affect. However, methods requiring daily reports of mood are limited by conceptual and practical difficulties in data analysis (Schwartz and Stone, 1998). We were motivated to develop the proposed model to relate continuous mood (affect) and discrete event patterns over time in order to diagnose minor or sub-threshold depression better. Our approach is to consider whether we can begin to develop a multivariate classification technique using affect and event reports to describe depressive disorder.

The Diagnostic and Statistical Manual—fourth edition (DSM-IV) of the American Psychiatric Association (2000) contains specific diagnostic criteria for psychiatric disorders such as major depression. For example, criteria for major depression require the presence of sad mood (dysphoria) or loss of interest in daily activities plus four of the following seven criteria: appetite disturbance, sleep disturbance, fatigue, psychomotor agitation or retardation, feelings of worthlessness, trouble concentrating, and thoughts of death or suicide. All symptoms must have been present for nearly every day for two weeks or more. However, persons with symptoms not meeting the full diagnostic criteria for major depression may still be at greater risk for physical illness, functional impairment, and death (Bruce et al., 1994; Gallo et al., 1997; Penninx et al., 1999, 1998).

Reports of depressive symptoms are usually evaluated by obtaining histories from patients or their informants in clinical or research assessments. Symptom counts based on recall may miss sporadic occurrences of sub-threshold symptoms that may be associated with significant disability or that might signal increased risk for the development of the full threshold diagnosis of major depression. For example, only 38% of persons in a community sample of adults recalled a lifetime history of dysphoric mood that they had reported 13 years earlier (Thompson et al., 2004).

The need to conceptualize depressive symptoms that do not reach full diagnostic threshold is particularly acute in the general medical setting against the background of medical conditions that accompany depressive symptoms (Bogner et al., 2002; Eaton et al., 1996; Waserthel-Smoller et al., 1996). In particular, depression in the context of cardiovascular disease (CVD) has been independently associated with poor outcomes. Employing modified criteria for major depression among persons who had recently sustained a myocardial infarction (MI), Frasure-Smith et al. (1993) found that six of 35 patients (17%) with major depression died after six months, while only six of 200 patients (3%) without major depression died—a difference that persisted after adjustment for functional status and previous MI. Based on a community survey of over 5000 persons, depressed persons with CVD at baseline were found to be about five times as likely to die from CVD than were persons with CVD without depression after six months of follow-up (Aromaa et al., 1994). In a four-year follow-up study of 450 persons with CVD (from a total sample of 2847 persons aged 55 to 85 years), Penninx found that major depression was associated with a three-fold increase in mortality compared to persons without depression, while minor depression was associated with a 1.6-fold increase (Penninx et al., 2001).

The issue of depression that does not meet the full criteria for major depression is pertinent to the study of depression following MI for several reasons. First, if we limit our attention to major depression, we are likely to ignore important phenomena that may be related to disability or other important patient-centered outcomes. From a public health perspective, there is good reason for casting the net wider, especially when considering depression among older adults (Gallo and Lebowitz, 1999). Minor depression has no standard
Bayesian model to identify trajectories of positive affect and negative events

operational criteria across studies: suggested research criteria in the DSM-IV are the same as for major depression except that the number of symptoms required for diagnosis is less than for major depression. Prevalence estimates of minor depression reveal a significant additional number of older adults would be identified with a depressive disorder (Tannock and Katona, 1995), consistent with studies finding that patients with ‘minor depression’ tend to be older and more physically ill (Blazer, 1991; Broadhead et al., 1990; Judd et al., 1994; Katon and Schulberg, 1992). Depressive symptoms that do not meet the full standard criteria for major depression, nevertheless, may increase risk for adverse outcomes and reactivity to negative external events over time.

Some evidence exists of negative correlations between positive affect measures and depression among the elderly (Beck et al., 2003; Lawton et al., 1996), although Wetherell et al. (2001) found that low positive affect may not be as specific to depression in older adults as in younger people. Thus, it would be of substantial clinical interest to determine if there are classes of subjects—non-depressed, mildly depressed, or more severely depressed—that could be determined by relating trajectories of mood measures on a background of time-invariant and time-varying characteristics to risk of full syndromal major depression. For example, Blazer et al. (1988) attempted to cluster depressive symptoms into ‘syndromes’ paralleling psychiatric disorders, finding five clusters (mild dysphoria, cognitive impairment, major depression, premenstrual syndrome, mixed anxiety–depression) sorted from over 80 reported individual symptoms and tests. Chen et al. (2000) and Eaton et al. (1989) employed latent class models to search for symptom clusters that might identify important subtypes of depression. Since these latent classes cannot be observed directly, it is natural to employ a mixture model in this setting. In this paper, we use the data on daily affect and events recorded after MI (Kumar et al., 1997) to extend the general growth mixture model (GGMM) (Muthén, 2001a,b; Nagin and Tremblay, 2001) to continuous (affect) and dichotomous (event) data over time. Our goal is to understand how mood and reactivity to negative events over time relate to diagnostic level depression. (Negative event reports are subjective: faced with the same events from ‘daily life’, people tend to interpret them in very different ways—first, deciding whether they are worth comment, and second, deciding whether to put a positive or negative cast on them.) The GGMM is an appropriate model to use because we can reduce daily affect and reported events into a small number of classes of subjects, which can then be associated with risk of major depression in the context of covariates measured at baseline. We implement the model in a Bayesian framework, which we believe is useful given the relatively small sample of patients and which allows us to consider a wide range of model selection and model checking procedures. We conclude with a discussion of the strengths and limitations of our approach, and directions for future applications.

2. A BAYESIAN GGMM WITH CONTINUOUS AND CATEGORICAL COVARIATES

Latent class analysis attempts data reduction by classifying subjects into one of $K$ unobserved classes based on observed data, where $K$ is fixed and known. Such latent classes are then identified in terms of their relationships with the observed variables. A simple latent class model relates the probability of a vector of (time-invariant) Bernoulli outcomes to membership in class $k$, $k = 1, \ldots, K$ (Clogg, 1995). The use of latent class models in the absence of agreed diagnostic criteria has precedence in the psychiatric literature (Chen et al., 2000; Eaton et al., 1989; Young, 1983).

Observed data for latent class models have traditionally corresponded to cross-sectional observed measurements (Goodman, 1974; Clogg and Goodman, 1985). More recently, latent class models have been extended to accommodate longitudinally observed data (Nagin, 1999; Muthén and Shedden, 1999; Roeder et al., 1999; Nagin and Tremblay, 2001; Muthén et al., 2002). In very general terms, these general growth mixture models or GGMMs assume that the trajectories of the observed data from $n$ subjects are driven by an underlying subject-level latent (unobserved) growth process. The mean structure for this
latent growth process itself depends on whether the subject belongs to one of \( K \) latent classes, where \( K \ll n \). This approach has several useful features. First, one focus of the analysis is on the estimation and interpretation of a small number \( K \) of latent classes, defined by the trajectories of latent growth curve processes, which, in turn, are defined by the observed data trajectories. Second, these latent classes can also be related to baseline covariates via regression models, further assisting in the interpretation of the latent classes. Third, estimates of the subject-level latent growth process can be obtained, as well as estimates of subject probability of membership in each of the estimated classes. This last point is somewhat subtle but important: while the model \textit{a priori} assumes that each subject is a member of a single class, the resulting posterior distribution for each subject is a mixture over all of the classes. If the classes are well defined, each subject will belong to a single class with high probability; and ill-defined single class, the resulting posterior distribution for each subject is a mixture over all of the classes. If the classes are well defined, each subject will belong to a single class with high probability; and ill-defined classes will leave many subjects with approximately \( 1/K \) probability of class membership. Such posterior probabilities of class membership, together with estimates of the latent growth curves themselves, may be useful to identify persons at risk for development of major depression or other outcomes.

To analyze the MI patient affect data, we propose the following GGMM:

\[
y_{it} | \eta_i \sim \text{ind } N(\Lambda \eta_i, \Sigma)
\]

\[
\eta_i | C_i \sim \text{ind } N_m(\beta_k, \Psi)
\]

\[
v_{it} | \tau_i \sim \text{BER} \left( e^{\Lambda \tau_i (\Lambda \eta_i + \delta_d + d_i)} / (1 + e^{\Lambda \tau_i (\Lambda \eta_i + \delta_d + d_i)}) \right)
\]

\[
\tau_i | C_i \sim \text{ind } \text{N}(\gamma_K, \Sigma).
\]

\[
C_i | x_i \sim \text{MULTI} (1; \pi(\delta_1, x_i), \ldots, \pi(\delta_K, x_i)),
\]

\[
\pi(\delta_k, x_i) = e^{\delta_k x_i} / \sum_{k=1}^{K} e^{\delta_k x_i}, \quad \delta_K \equiv 0
\]

The \( y_{it} \) are the positive affect scores and the \( v_{it} \) are indicators of presence or absence of daily perceived negative events for the \( i \)th subject at day \( t \), \( t = 1, \ldots, n_i \), \( i = 1, \ldots, n \). \( C_i \) gives the latent class membership status: conditional on \( C_i \), both the latent \( \eta_i \) and the \( \tau_i \) are independent. This extension of the latent class growth curve model lets us consider how the trajectories of the affect and event measures may simultaneously be related to the underlying latent class membership, which, in turn, is related to a baseline measure of major depression. For ease of estimation, we rewrite as \( \eta_i = \beta_k + \delta_d \) and \( \tau_i = \gamma_k + d_i \), \( \beta_k \sim N(0, \Psi) \), \( d_i \sim N(0, \Sigma) \).

Since the \( \eta_i \) are unobserved, \( \Lambda \eta_i \) (given by the \( m \times n_i \) matrix consisting of the stacked \( m \times 1 \) row vectors \( \Lambda \eta_i \)) and \( \Psi \) are aliased, that is, there is a non-unique set of values that will provide an equally good fit to the data. If \( \Psi \) is fixed as an identity matrix, a common \( \Lambda \eta_i \equiv \Lambda \gamma \) could be estimated and interpreted as a matrix of factor loadings. Alternatively, we fix \( \Lambda \eta_i \) as in a polynomial function of known time points of observations, allowing \( \Psi \) to be estimated. As with \( \Lambda \eta_i \), we consider \( \Lambda \gamma \) to be an \( l \times n_i \) matrix consisting of the stacked \( l \times 1 \) row vectors \( \Lambda \gamma_i \), of known time points of observations, which allows \( \Sigma \) to be estimated. We focus on this ‘growth curve’ alternative throughout the remainder of this article. In particular, we consider a quadratic growth curve for the positive affect scores and a linear growth curve for the log-odds of reporting a negative event. This model posits a quadratic subject-level growth curve \( \eta_i \) that relates the observed continuous data \( y_i \) to the mean growth curve for class \( k \) parameterized by \( \beta_k \) (Verbeke and Lesaffre, 1996), and extends this model to include a linear subject-level growth curve \( \tau_i \) that relates the log-odds of the observed dichotomous data \( v_i \) to the mean growth curve for class \( k \) parameterized by \( \gamma_k \). The elements of \( \delta \) carry the information about the probability of class membership present in \( x \); in our application, we consider \( x \) to include an intercept and an indicator of baseline depression. Figure 1 shows a directed graph of the hierarchical model given by (2.1).
This model is based on the GGMM model given in Muthén and Shedden (1999) and Muthén et al. (2002) but differs in two important ways. First, rather than considering a single baseline or outcome categorical variable, (2.1) accommodates a vector of $v_{it}$ categorical longitudinal outcomes for each subject. In this way, our model is similar to the ‘two-process’ models of Nagin (1999), which consider trajectories of both Poisson and censored-normal outcomes linked by a common latent class assignment. Second, we estimate (2.1) and perform model validation with Bayesian techniques. We use a Bayesian approach for three reasons. The hierarchical nature of the latent class growth curve models suggests extending it to include prior distributions on the unknown parameters—such an approach has been taken by Garrett and Zeger (2000) for a simpler model including only baseline Bernoulli outcomes. Second, the EM procedure required to estimate (2.1) in a frequentist setting was problematic because the multimodality of the mixture likelihood meant that different startpoints yielded different convergence points, leaving us uncertain if we
had reached all of the modes. Finally, while we have a large number of observations per subject, our total number of subjects (we consider a dataset with \( n = 35 \)) is small compared with typical latent class growth curve models. Our use of Bayesian methods avoids the need to rely on large-sample maximum likelihood or bootstrap approximations for inference. This approach also allows for a more complete description of the parameters of interest via examination of posterior distributions and incorporation of uncertainty in class assignment into inference about these parameters.

2.1 Prior distributions

For the class membership parameters, we consider proper priors of the form (Garrett and Zeger, 2000)

\[
\delta_k \sim N(0, 9/4 I).
\]  

(2.2)

After the logit transformation, this provides priors for \( \pi_k \) that are centered at \( 1/K \) and are relatively flat elsewhere, going to 0 near 0 and 1 (see Figure 1 in Garrett and Zeger, 2000).

For the growth curve parameters \( \beta_k \) and \( \gamma_k \), informative priors are needed to account for the fact that the probability for given class \( k \) may be small with a non-trivial posterior probability. If this is the case, the posterior of \( \beta_k \) or \( \gamma_k \) will be improper unless the prior is proper. In particular, we assume independent multivariate normal distributions centered at the linear regression estimate \( \beta^* \) of \( y \) on \( \Lambda_1 y \) and at the logistic regression estimate \( \gamma^* \) of \( v \) on \( \Lambda_w \), with covariances equal to \( n \) times the covariance estimates of \( \beta^* \) and \( \gamma^* \) obtained as the inverse of the expected information matrix:

\[
P(\beta_k) \overset{\text{ind}}{\sim} N(\beta^*, \Psi_0)  
P(\gamma_k) \overset{\text{ind}}{\sim} N(\gamma^*, \Sigma_0).
\]  

(2.3)

This data-based prior tends to force the posterior trajectories of the classes to be more equal to each other than a purely non-informative prior would but scaling the prior variance by the sample size ensures that this effect is very weak unless the sample is quite small.

Finally, because of the relatively large amount of data for the variance parameters that are fixed across the classes (in contrast to the class-specific \( \beta_k \) and \( \gamma_k \) parameters), we assume independent inverse-gamma and inverse-Wishart priors:

\[
p(\sigma_1^2, \ldots, \sigma_p^2, \Psi, \Sigma) \propto \prod_i (\sigma_i^2)^{-(a+1)}e^{-b/\sigma_i^2} \times |\Psi|^{-(v_1+m+1)/2} \exp[-1/2tr(S_1\Psi^{-1})] |\Sigma|^{-(v_2+l+1)/2} \exp[-1/2tr(S_2\Sigma^{-1})].
\]  

(2.4)

In the following analyses, we set \( a = b = 0 \), yielding the standard non-informative prior on the residual variances, and \( v_1 = m, v_2 = l, S_1 = \text{diag}(0.01) \) and \( S_2 = \text{diag}(0.001) \), weakly informative inverse-Wishart priors that ensure a proper posterior.
2.2 Obtaining the posterior distribution

We consider the full posterior distribution given by

\[
p(\eta, \tau, \beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_K, \delta_1, \ldots, \delta_K, \sigma_1^2, \ldots, \sigma_n^2; \Psi, \Sigma | y,v,x) \]

\[
\propto \prod_{i=1}^{n} \left[ f(y_i, v_i | \eta_i, \tau_i, C_i = k) p(\eta_i, \tau_i | C_i = k) p(C_i = k | x_i) \right]
\times p(\beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_K, \delta_1, \ldots, \delta_K, \sigma_1^2, \ldots, \sigma_n^2; \Psi, \Sigma)
\]

\[
= \prod_{i=1}^{n} \prod_{l=1}^{p} \left[ f(y_{il} | \eta_l) f(v_{il} | \tau_l) \right] p(\eta_l | C_i = k) p(\tau_l | C_i = k) p(C_i = k | x_i)
\]

\[
\times p(\sigma_1^2, \ldots, \sigma_n^2; \Psi, \Sigma) \prod_{k=1}^{K} \left[ p(\beta_k) p(\gamma_k) \right] \prod_{k=1}^{K-1} p(\delta_k)
\]

(2.5)

where the last equality follows from four conditional assumptions of independence:

1. independence within the elements of the continuous outcomes \(y_i\) and the categorical outcomes \(v_i\) conditional on the latent subject-level growth curves \(\eta_i\) and \(\tau_i\);
2. independence between \(y_i\) and \(v_i\) conditional on the latent subject-level growth curves \(\eta_i\) and \(\tau_i\);
3. independence between the latent growth curves \(\eta_i\) and \(\tau_i\) conditional on latent class status \(C_i\); and
4. independence among the growth curve parameter priors.

The right-hand side of the last equality in (2.5) is given by (2.1)–(2.4).

The posterior distribution is obtained using a Markov Chain Monte Carlo (MCMC) approach that uses Gibbs sampling (Gelfand and Smith, 1990; Gelman and Rubin, 1992). In brief, Gibbs sampling obtains draws from a joint distribution of \(p(\theta | \text{data})\) for \(\theta = \{\theta_1, \ldots, \theta_q\}\) by initializing \(\theta\) at some reasonable \(\theta^{(0)}\) and drawing \(\theta_1^{(1)}\) from \(p(\theta_1 | \theta_2^{(0)}, \ldots, \theta_q^{(0)}, \text{data})\), \(\theta_2^{(1)}\) from \(p(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_q^{(0)}, \text{data})\), and so forth. As \(T \to \infty\), \(\theta^{(T)} \sim p(\theta_1, \ldots, \theta_q | \text{data})\). A detailed description of the MCMC algorithm used to obtain simulations from the posterior is given in the Appendix.

A difficulty with MCMC estimation of posterior distributions in latent class models is the 'label-switching' problem that results from the fact that permutations of the class assignments are not identifiable since the likelihood is unchanged under these permutations (Redner and Walker, 1984). Consequently, an assignment to 'Class \(k\)' during one part of the Gibbs chain may have the same meaning in terms of the underlying model structure as an assignment to 'Class \(k'\' during another part of the Gibbs chain, with the assignment probabilities to a given placeholder converging to \(1/K\) as the number of simulations from the posterior increase to infinity. To 'untangle' the assignments, we use the decision-theoretic post-processing approach described by Stephens (2000). In this approach, all possible permutations of class assignments are considered at each iteration of the Gibbs sampler and the permutation of the class assignment that chosen maximizes the posterior probability that the labeling of classes is consistent with the previous assignments.

We obtained draws from five chains of 5000 draws each, after discarding 100 for burn-in. After implementing the Stephens label-switching algorithm within each chain, we obtained the Gelman–Rubin statistic \(\hat{R}\), an adjusted ratio of between- and within-sequence variability to within-sequence variability (Gelman et al., 1995, p. 331–333). At convergence, \(\hat{R} = 1\): generally \(\sqrt{\hat{R}} = 1.2\) is considered sufficient for convergence. Here, \(\max \sqrt{\hat{R}} = 1.03\), even within the highly multi-modal subject-level probabilities of class membership.

The R software (http://cran.r-project.org/) for fitting this model is available at http://www.cceb.med.upenn.edu/elliott/.
3. Application to Affect and Event Data from Subjects Post-Myocardial Infarction

Affect score and presence of positive, negative, and neutral events were observed in 35 patients who had experienced an MI within the past year and were in treatment at a University of Pennsylvania cardiology clinic. These patients were recruited to participate in a pharmacological and neuroimaging study of elderly patients and included both subjects who met SCID (Structural Clinical Interview DSM-IV) criteria for threshold minor depressive disorder and those without depression (Kumar et al., 1997). Affect scores and event indicators were collected for up to 35 consecutive days; and complete data were available for 20 subjects, with a mean of 31.5 days of observed data. Subjects reported negative events on 45% of days, with individual subjects reporting as few as 8% and as many as 94% of their days having negative events. Positive affect was measured by asking subjects whether they were feeling ‘energetic’, ‘warm toward others’, ‘interested’, ‘happy’, or ‘content’ at that moment, with a 1 (not at all) to 5 (extremely) rating scale for each item. Thus, positive affect scores ranged from 5 to 25, with a mean of 14.8 and a standard deviation of 4.6; and the mean within-subjects positive affect scores ranged from 7.2 to 23.0. Figure 2 shows the positive affect scores and negative event reports for all 35 subjects.

Preliminary analysis using standard linear and generalized linear mixed models indicated that fixed effects were statistically significant up to the quadratic term for the positive affect growth curves and up
Bayesian model to identify trajectories of positive affect and negative events

127

to a linear term for the log-odds of reporting a negative event; hence, in (2.1),
\[
\Lambda_{y_i} = \begin{bmatrix}
1 & (t_i - T_i) & (t_i - T_i)^2 \\
\vdots & \vdots & \vdots \\
1 & (t_{n_i} - T_i) & (t_{n_i} - T_i)^2
\end{bmatrix}
\]
and
\[
\Lambda_{y_i} = \begin{bmatrix}
1 & (t_i - T_i) \\
\vdots & \vdots \\
1 & (t_{n_i} - T_i)
\end{bmatrix},
\]
where \(T_i\) is the mean reporting time for the \(i\)-th subject. Each subject also has unobserved random effects corresponding to intercept, slope, and quadratic linear regression of their affect growth curve parameters given by \(\eta_i = (\eta_{i0} \eta_{i1} \eta_{i2})'\) and unobserved random effects corresponding to intercept and slope logistic regression parameters of their event growth curve given by \(\gamma_i = (\gamma_{i0} \gamma_{i1})'\) (thus, \(m = 3\) and \(l = 2\)). Additionally, we considered as major or threshold minor depression (6% of subjects) as the key predictor of class status; thus, \(x_i\) in (2.1) is a two-component vector for each subject, consisting of an intercept and an indicator of depression status.

Preliminary analysis using standard linear and generalized linear mixed models, as well as runs of the Gibbs sampler, indicated that the variance \(\psi_{22}\) of the quadratic random effect \(\eta_{i2}\) was near zero, suggesting that a random-slope intercept model was sufficient to capture subject-level heterogeneity in the positive affect growth curves; hence,
\[
\Psi = \begin{bmatrix}
\Psi_{00} & \Psi_{01} & 0 \\
\Psi_{01} & \Psi_{11} & 0 \\
0 & 0 & 0
\end{bmatrix}.
\]
Similarly, a random intercept model was sufficient to capture subject-level heterogeneity in the negative event growth curves; and, hence, \(\Sigma = \begin{bmatrix} \Sigma_{00} & 0 \\ 0 & 0 \end{bmatrix} \). We restrict our analysis to two- and three-class models, given the limited number of subjects (\(n = 35\)).

3.1 Assuming one class

Assuming one class is equivalent to fitting separate linear mixed models for the positive affect data and generalized linear mixed models for the negative event data: Figure 3 shows that, averaged across all 35 subjects, positive affect declines from approximately 16 to 14 between Day 1 and 20, then stabilizes. On average, the probability of reporting a negative event declines steadily from about 0.5 to 0.35.

3.2 Assuming two classes

Figure 4 shows the growth curve trajectories for each of the two classes. Class 1 consists of those who have relatively constant positive affect scores and declining probabilities of negative events, while Class 2 consists of those whose affect scores begin slightly higher than those of Class 1 but drop below and level off, and whose probability of negative events remains high over the observed time period. We term Class 1 subjects ‘optimists’ and Class 2 subjects ‘pessimists’, as a rough interpretation of the growth curves.

Figure 5 considers the posterior distribution of \(\pi(x_k)\), the probability of being a member of a class as a function of either being depressed (\(\pi(\text{depressed})_k = e^{\delta_{k1} + \delta_{k2}} / \sum_k e^{\delta_{k1} + \delta_{k2}}\)) or non-depressed (\(\pi(\text{non-depressed})_k = e^{\delta_{k1}} / \sum_k e^{\delta_{k1}}\)). Depressed subjects have a posterior mode of 90% of being ‘pessimists’, compared with a posterior mode of 68% of non-depressed subjects—though the 95% posterior predictive intervals are quite wide, ranging from 12–98% among the depressed and 22–88% among the non-depressed.

Note that this posterior distribution assumes knowledge of the baseline covariates only: once affect and event indicators are observed, subject-level posterior probabilities of belonging to the ‘optimist’ class
can be determined as the posterior distribution of $\pi^*(y_i, v_i, x_i) = \pi_{i1}/(\pi_{i1} + \pi_{i2})$, where

$$\pi_{ik} = \prod_t \phi(y_{it}; \Lambda_{yi}[\beta_k + b_i], \sigma_i^2) \exp(\Lambda_{yi}[\gamma_k + d_i])^{v_{it}} \left[ e^{\delta_k x_i} / \left( \sum_{k=1}^K e^{\delta_k x_i} \right) \right]$$

and

$$\phi(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

is the PDF of a normal distribution with mean $\mu$ and variance $\sigma^2$. Figure 6 shows the posterior mean subject-level growth curves for two subjects with high posterior probability of membership in Class 1 and Class 2, respectively.

Ideally, all subjects in the two-class model would belong to either the pessimist or optimist class with a high posterior probability $\pi^*(y_i, v_i, x_i)$, $j = 1, 2$. While 11 subjects have a very high posterior probability of class membership (greater than 0.8), eight subjects have posterior probabilities between 0.4 and 0.6, indicating that, for these subjects, the two-class model is not effective at classifying their growth curves. Hence, we consider a three-class model.

### 3.3 Assuming three classes

Figure 7 shows the growth curve trajectories for each of the three classes. Classes 1 and 2 show initially high but declining positive affect measures, similar to Class 1 in the two-class model; however, Class 2 shows a more rapid decline than Class 1 in the probability of negative events being reported. Class 3 shows fairly stable positive affect and a stable, intermediate risk of reporting negative events. For rough interpretive labels, we term Class 1 subjects ‘pessimists: stable’, Class 2 subjects ‘pessimists: declining’, and Class 3 subjects ‘optimists’.
Figure 8 suggests that there might be little information in the data to determine the relationship between depression and classification, suggesting that the model might be overfitting by including the covariate (see Section 3.5.2 later). Depressed subjects did appear somewhat less likely than non-depressed subjects to belong to the less pessimistic Class 2 and Class 3 groups (posterior modes of 4% and 6% among the depressed versus 17% and 20%, respectively, among the non-depressed) but there was little information in the data to relate depression to the most pessimistic Class 1 grouping.

Considering the posterior probabilities of class membership, there are no subjects that belong to ‘pessimists: declining reported negative events’ class with a posterior probability greater than 0.5. Seventeen subjects belong to ‘pessimists: stable reported negative events’ with a posterior probability greater than 0.5 but only two with a posterior probability greater than 0.8. Two subjects belong to ‘optimist’ with a very high (>0.9) posterior probability, but 30 belong to this class with a posterior probability of <0.5. Thirteen subjects remain not well classified, with mean subject-level posterior probabilities of class membership below 0.5 for all three classes.
Fig. 5. Two-class model: posterior distribution of class probabilities given depression status: full curve (---) for depressed and dashes (---) for non-depressed. Class 1 = ‘optimist’, Class 2 = ‘pessimist’.

3.4 Modeling the continuous and discrete outcomes separately

As a confirmatory analysis, we consider the results for the two- and three-class models, fitting the class structure using only the continuous positive affect measures or only the categorical negative event reports (Nagin and Tremblay, 2001). Figure 9 shows that, for the two-class model, the growth curves for the positive affect measures are distinct and similar to what was obtained from the joint estimation procedure; the negative events rates are somewhat less distinct than the joint model, although they do distinguish between a slow and more rapid decline in negative event reporting. The posterior mode of the probability of class membership in Class 1 is 0.30 for the non-depressed and 0.08 for the depressed, when positive affect is used to form the classes, and 0.26 for the non-depressed and 0.06 for the depressed, when negative event reporting is used to form the classes.

For the three-class model, the growth curve trajectories determined using the positive affect measures alone are similar to those obtained using the joint modeling (see Figure 10). However, the negative event reporting curves are far less similar. In contrast to fitting separate models using only continuous positive affect or only categorical negative event reports, the relationship between the two is made more explicit with a joint modeling strategy.

3.5 Model selection

The latent class growth curve model assumes that the number of classes $K$ is fixed and known. The choice of $K$ is problematic in both frequentist and Bayesian settings; however, the Bayesian approach allows us to employ several model checking and model selection techniques considered in the literature as well as explore others. We consider three approaches. First, we consider the ‘deviance information criterion’ extension of AIC- and BIC-type measures (Spiegelhalter et al., 2002). Second, by comparing the prior and posterior distributions for the $\delta$ parameters governing baseline class probabilities, we can determine if the number of classes is too large, leading to identifiability problems (Garrett and Zeger, 2000). Finally, to determine if the number of classes is too small, we use posterior predictive checks (Gelman et al., 1996): replicated data generated under the model is compared to the observed data for evidence of model failure.

Of course, in choosing the number of classes, interpretability counts as well. A model that presents latent classes with a clear and simple interpretation might be preferred to one that provides somewhat better fit but does not yield as much insight.
Bayesian model to identify trajectories of positive affect and negative events

3.5.1 Likelihood-based summary measures. Table 1 gives the results of $-2\log$-likelihood evaluated at the posterior median of the parameter estimates. The log-likelihood for the $K$-class model is given by

$$L_K(\theta) = \sum_i l_i(\theta_i)_K$$

where

$$l_i(\theta_i)_K = \log \sum_{k=1}^K \pi(\delta_k, x_i) \left\{ \phi(y_i; \Lambda_{y_i}, \beta_k, \sigma^2 1 + \Lambda_{y_i}, \Psi \Lambda'_{y_i}) \int \prod_{t=1}^{n_i} \left( \frac{e^{\Lambda_{y_t}(\gamma_{t, k} + \delta_i)_{y_t}}}{1 + e^{\Lambda_{y_t}(\gamma_{t, k} + \delta_i)_{y_t}}} \right) \phi(d_t, 0, \Sigma) \, dd_t \right\}$$

(3.1)

where $\theta_i = (\beta, \gamma, \sigma^2, \Psi, \Sigma, \delta)$ and the integral over the random effect of the Bernoulli outcome is estimated via Gaussian quadrature.

In general, larger models give better estimates of model fit: standard penalties such as AIC (Akaike, 1978) or BIC (Schwarz, 1978) assume that the number of parameters is a known quantity but, in a hierarchical Bayesian framework, the number of effective parameters may be unclear. The random
effects associated with each subject may ‘count’ as approximately one parameter if the between-variance
estimates are large (small degree of shrinkage) and as nearly zero parameters if the between-variance
estimates are small (large degree of shrinkage). To estimate the effective number of parameters, we use
the ‘deviance information criterion’ method of Spiegelhalter et al. (2002). In this approach, the deviance
$D(\theta) = -2L_K(\theta) + 2\log(f(y))$ for the $K$-class model can be thought of as the Kullback–Leibler
information with ‘excess’ information from overfitting given by $f(y) = \sum_i \int f(y_i|\theta)p(\theta)\,d\theta$. The
effective number of parameters $p_K$ is then estimated by the ‘mean of the deviance minus the deviance of
the mean’ or

$$p_K = E_{\theta|y,v}(D(\theta)_K) - D(E_{\theta|y,v}(\theta))_K$$  \hspace{1cm} (3.2)

(the second term in $D(\theta)$ is a function of the data only and can be ignored when computing $p_K$). The DIC

Fig. 7. Three-class model: posterior median of growth curve trajectory for positive affect and probability of negative
event, by class. Broken curves provide 95% posterior predictive intervals. Class 1 = ‘pessimists: stable’, Class 2 =
‘pessimists: declining’, Class 3 = ‘optimist’.
Bayesian model to identify trajectories of positive affect and negative events

Fig. 8. Three-class model: posterior distribution of class probabilities given depression status: full curve (—) for depressed and broken curves (---) for non-depressed. Class 1 = 'pessimists: stable', Class 2 = 'pessimists: declining', Class 3 = 'optimist'.

measure is given by $D(E_{\theta|y,v}(\theta))_K + 2p_K$. By the DIC criterion, the three-class model is favored slightly over the two-class model (see Table 1).

The DIC measure has been criticized, among other reasons, on the grounds that it is sensitive to the parameterization of the model (e.g. formulating the $t$-distribution directly versus formulating it as a gamma mixture of normal distributions: see Smith, Gelfand and Trevisani, in discussion of Spiegelhalter et al., 2002); that its derivation lacks theoretical justification (Plummer, Robert and Titterington, in discussion of Spiegelhalter et al., 2002); and that, in mixture models such as this one, the penalty for the increased number of parameters in more complex models does not function correctly, yielding DIC measures that may favor overly complex models (Richardson, in discussion of Spiegelhalter et al., 2002). There is some evidence of the latter in our results here, given the relatively weak identification of the higher-class models by the criterion of Garrett and Zeger (2000), as we see later.

An alternative Bayesian criterion for choosing between models of class size $K$ and $K'$ are Bayes factors. This approach can be computationally complex, as well as sensitive to the choice of priors in weakly identified models. We defer exploration of Bayes factors in this paper.

3.5.2 Comparing priors and posteriors. We next present an alternative measure to consider model fit. With proper priors, posteriors are always proper and, thus, technically identified. However, if there are no data available to distinguish parameters, the posteriors will match the priors; hence, the ratio of the prior and posterior densities can be considered as a measure of ‘Bayesian identifiability’ (Garrett and Zeger, 2000). We consider these ratios by plotting priors and posteriors graphically in Figures 11 and 12.

Fixing $K = 2$, the posterior of the logit parameters governing the relationship between the latent class assignment and the intercept and baseline covariate of depression are shown in Figure 11. The average overlap between prior and posterior distributions for the intercept is small, indicating that the two-class model is well identified (Garrett and Zeger, 2000) and, hence, relatively insensitive to prior assumptions. However, there is substantial overlap between the prior and posterior for the depression indicator, indicating that, while there are some data to estimate a negative association between membership in the optimist class and depression status, much of the information about this parameter is being supplied from the prior assumption. (The median of $\delta_{12}$ is $-0.52$, with a 95% posterior predictive interval [PPI] of $(-3.25, 2.14)$, compared with the prior median of 0 and 95% predictive interval of $(-2.94, 2.94)$.)

Figure 12 shows the equivalent posterior-versus-prior plots for $K = 3$. Note that the overlap between the prior and posterior for the intercept is greater than in the two-class model, particularly for Class 2.
Fig. 9. Two-class model: posterior median of growth curve trajectory for positive affect and probability of negative event, by class, estimating the classes using positive affect alone or negative event alone. Broken curves provide 95% posterior predictive intervals.

(pessimists with declining reported negative event rates). The ability to link the depression measures with the class membership has virtually disappeared, as the near overlap between the prior and posterior for $\delta_{12}$ and $\delta_{22}$ indicates. This is evidence that three classes may not be well identified from the data.

### 3.6 Model checking

In addition to class size, we need to consider whether the independent normal and Bernoulli assumptions conditional on the unobserved random effects and latent classes are sufficient. Posterior predictive distributions (Gelman et al., 1995, 1996) can be used to test both the distributional and latent class number assumptions of the model. Briefly, classic $p$-values represent the probability under the model at some fixed parameter value $\theta$ that the observed statistics $S(y)$ will be less than (or greater than) the values of the statistic that would be seen in repeated experiments: $P(S(y) \leq S(y^{\text{rep}})|\theta)$. The posterior predictive distribution (PPD) $p$-value represents the probability that the observed statistic (which can be a function of
Fig. 10. Three-class model: posterior median of growth curve trajectory for positive affect and probability of negative event, by class, estimating the classes using positive affect alone or negative event alone. Broken curves provide 95% posterior predictive intervals.

Table 1. $-2 \times \log$-likelihood for k-class models evaluated at the posterior mean of their parameter estimates, together with DIC penalty. Best fit is determined by smallest value

<table>
<thead>
<tr>
<th>Number of classes $k$</th>
<th>$-2 \times \text{LL}$</th>
<th>$p_k$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>6670.75</td>
<td>35.91</td>
<td>6742.47</td>
</tr>
<tr>
<td>Two</td>
<td>6655.54</td>
<td>40.40</td>
<td>6736.54</td>
</tr>
<tr>
<td>Three</td>
<td>6648.16</td>
<td>43.08</td>
<td>6734.34</td>
</tr>
</tbody>
</table>

both the data $y$ and the parameter $\theta$) is more extreme than replicated statistic, conditional on the observed data: $P(S(y, \theta) \leq S(y^{\text{rep}}, \theta)|y)$. PPD $p$-values can be obtained from the draws of $\theta$ generated by the Gibbs sampler; and $y^{\text{rep}}$ can be drawn from $f(y|\theta^{\text{rep}})$, and $S(y^{\text{rep}}, \theta^{\text{rep}})$ compared with $S(y, \theta^{\text{rep}})$.

We consider two posterior predictive statistics to test the distributional assumptions of the time-varying covariates. First, for the quasi-continuous positive affect data, we consider whether the assumption of normality of the positive affect score conditional on the latent subject-level growth curve is reasonable using the chi-square-type statistic

$$S_1 = N^{-1} \sum_i \sum_t \frac{(y_{it} - \mu_{it})^2}{\sigma_i^2} \quad (3.3)$$
Fig. 11. Two class model: prior (broken curve) and posterior (full curve) distribution of $\delta_1$: $\delta_{11}$ is the log relative risk of belonging to Class 1 rather than Class 2 when $x_i = 0$ (non-depressed); $\delta_{12}$ is the log of the ratio of relative risks of belonging to Class 1 rather than Class 2 for those suffering major depression versus others. Class 1 = ‘optimist’, Class 2 = ‘pessimist’.

where $\mu_{it} = E(y_{it}|b_i) = \Lambda'_{t \beta_k} \pi_{ik} \beta_k + b_i$ and $N = \sum_i n_i$. Note that the replicated data will follow a $\chi^2_N$ distribution. Second, for the Bernoulli event data, we consider whether the latent subject-level growth curve model for the logit of the negative event indicator probabilities is reasonable using the subject-level statistic

$$S_{2i} = \sum_i v_{it}.$$  \hfill (3.4)

Figure 13 shows the PPD of observed versus replicated $\chi^2$ values given by (3.6) for the one-, two-, and three-class model. There is no evidence of substantial model failure, although the slight underdispersion of the observed $S_1$ hints at the mild overfitting due to the fact that the positive affect scores are not truly continuous. The PPD $p$-values for the one-, two-, and three-class models are 0.30, 0.24, and 0.24, respectively. The somewhat lower $p$-values for the two- and three-class models may reflect the better discrimination in the growth curves, which may highlight, to some degree, the discrete nature of the affect measures.

Next, $\sum_{i=1}^{\text{rep}} I(\sum_{i=1}^{\text{rep}} v_{it}^{\text{rep}} > \sum_{i=1}^{\text{obs}} v_{it}^{\text{obs}}) / \sum_{i=1}^{\text{rep}} 1$ ranges across subject index $i$ from 0.13 to 0.62 for the one-class model, 0.11 to 0.62 for the two-class model, and from 0.09 to 0.63 for the three-class model. The most extreme case is a subject who reported negative events on 34 of 35 days; nonetheless, the model fits this rather extreme situation rather well, with the median of $\sum_{i=1}^{\text{rep}} v_{it}^{\text{rep}}$ equal to 33 under all three models, again indicating that the independent Bernoulli assumption conditional on the latent logistic growth curve is reasonable for the event data, regardless of class size.

4. DISCUSSION

This article details an extension of latent growth models that incorporates both continuous and categorical longitudinal outcomes, and allows for latent growth curve estimation among an expanded set of models. In addition, these models are described in a Bayesian framework that ensures against over- and underfitting through judicious use of priors and careful inspection of resulting posteriors. Use of Markov
Bayesian model to identify trajectories of positive affect and negative events

Fig. 12. Three-class model: prior (broken curve) and posterior (full curve) distribution of $\delta_1$ and $\delta_2$: $\delta_1$ is the log relative risk of belonging to Class 1 rather than Class 3 when $x_i = 0$ (non-depressed); $\delta_2$ is the log of the ratio of relative risks of belonging to Class 1 rather than Class 3 for those suffering major depression versus others. $\delta_1$ is the log relative risk of belonging to Class 2 rather than Class 3 when $x_i = 0$ (non-depressed); $\delta_2$ is the log of the ratio of relative risks of belonging to Class 2 rather than Class 3 for those suffering major depression versus others. Class 1 = ‘Pessimists: stable’, Class 2 = ‘Pessimists: declining’, Class 3 = ‘optimist’.

Chain Monte Carlo also provides draws from the full joint posterior so that a variety of interesting joint and marginal posterior distributions of interest among subsets of parameters can be considered.

It appears that, for the MI patients considered, the two-class model does pick out some hitherto unobserved structure in the positive affect and negative event data in this sample: an ‘optimist’ class with stable positive affect and declining perceived negative events; and a ‘pessimist’ class with declining positive affect and continuing perceived negative events. Depressed subjects had a 92% chance of belonging to a latent class associated with decreasing positive affect and persistent reporting of negative events, compared with 62% among non-depressed subjects. A three-class model appears to divide the ‘pessimist’ class into two sub-classes, one with high and stable levels of perceived negative events in their life, the other with declining rates of negative events in their life. The relationship between the three-class...
model and depression is difficult to discern because of the small number of depressed subjects in the sample.

One important clinically-relevant implication of our application of latent class growth curve modeling is that we might be able to identify persons post-MI who are most at risk of developing criterion-based major or threshold minor depression. In other words, we have linked—albeit weakly—a dimensional measure of affect over time and daily life events to standard categorical diagnoses for major depression and threshold minor depression. In larger samples, with more complete information on course-modifying characteristics, the clinical application might be to identify persons who require more intensive follow-up post-MI or for whom interventions could be targeted to prevent the onset of major depression. From the point of view of intervention and evaluation, we could introduce a term into the model representing treatment-condition assignment to study the role that participation in one or another treatment condition has on decreasing the risk for the development of depression. Also important to note is that the distinctions between these classes is apparent in the latent class models that include both affect and event trajectories but is not as apparent, particularly with regard to the event trajectories, in the analyses that considered only affects or events.
The analyses presented here have a number of limitations, primarily due to the relatively small sample. With only 35 subjects, it is difficult to consider models with large numbers of classes. Our analysis is primarily exploratory—are there a small number of ‘types’ of affect and event trajectories into which we can classify subjects that we also might be able to relate to subject-level covariates? If a confirmatory model is to be explored, a larger sample will be required to detect less-well-represented classes. An additional limitation is that we do not have any measures of cardiovascular health or function that might be related to the development of depression or negative events.

A number of extensions of this analysis are also possible, although the limited sample size may prevent their application in this dataset. First, we could allow the covariance matrices for the subject-level mean positive affect trajectory parameters $\Psi$ and negative event trajectories parameters $\Sigma$ to differ by class, as in Muthén et al. (2002). Second, we could allow correlation between the mean affect and event trajectories, even conditional on class assignment, that is, assume

$$\begin{align*}
\begin{pmatrix}
\eta_i \\
\tau_i
\end{pmatrix} \mid C_i = k &\sim N_{m+1} \left(\begin{pmatrix} \beta_k \\ \gamma_k \end{pmatrix}, \begin{pmatrix} \Psi & \Omega \\ \Omega & \Sigma \end{pmatrix}\right).
\end{align*}$$

(In (2.1), $\Omega \equiv 0$.) Third, non-normal forms of $y_i$ can be considered, as in Roeder et al. (1999), or interactions between class membership parameters governing the growth curve $\beta_k$ and $\gamma_k$ and covariate effects $x_i$, as in Muthén and Shedden (1999) or Muthén et al. (2002). Finally, Figure 1 suggests that subjects have differing degrees of within-subject variability with respect to positive affect: some subjects are highly autocorrelated, while others show large day-to-day random variation. We are currently considering models that focus on the residual variance after fitting a non-parametric smoothing spline curve to the continuous affect data; these could be viewed as a means of distinguishing between these subjects, which might have differing subsyndromal depression risks.

Despite the limitations related to the specific dataset we have analyzed, our use of the general growth mixture model deserves attention. Previous attempts to link affect and event reporting data over time to who does and does not meet criteria for depression have not been entirely successful and the GGMM approach attempts data reduction in a fashion that is consistent with the clinician’s tendency to classify patients into disease categories. The ultimate goal, of course, is to determine which subjects are at risk of negative health consequences due to minor depression and to target available resources better to prevent these negative consequences from occurring. The general growth mixture model, as implemented here, can provide a useful way to conceptualize the link between symptoms, life events, and outcomes.

ACKNOWLEDGEMENTS

This research was supported in part by National Institute of Mental Health grants P30-MH2129 and R01-MH-61892. The authors would like to thank Edmund Weisburg at the University of Pennsylvania for editorial assistance, as well as the two anonymous reviewers, the Associate Editor, and the Editors, whose comments and suggestions helped to improve this paper.

APPENDIX A

Conditional draws of the Gibbs sampler

Here we describe the conditional draws of the Gibbs sampler. Where conditional draws cannot be made directly from a known distribution, a variation of a sampling-importance-resampling (SIR) algorithm (Raghunathan, 1994) is used:
1. Draws are made from a dominating distribution \( N(\hat{\xi}_j, C\hat{V}_j) \), where \( \hat{\xi}_j \) is the mode of the conditional posterior distribution \( p(\xi_j | \xi_1, \ldots, \xi_{j-1}, \xi_{j+1}, \ldots, J, \text{data}) \equiv p(\xi_j | \text{rest}) \) obtained, e.g. via a Newton–Raphson algorithm, \( \hat{V}_j \) is the asymptotic covariance of the model estimator given by the negative of the inverse of the Hessian matrix of second derivatives of \( \log p(\xi_j | \text{rest}) \) evaluated at \( \hat{\xi}_j \), and \( C \) is a rescaling constant chosen to help ensure domination.

2. A single draw of \( \xi^d_j \) is then made from \( \xi^d_j, \ldots, \xi^D_j \) with probability proportional to \( w_d = p(\xi^d_j | \text{rest})/\hat{V}^{-1}_j (\hat{\xi}_j - \xi^d_j)/\sqrt{\hat{V}_j} \), where \( \phi(\cdot) \) is the standard normal PDF.

The conditional draws are made as follows (here \( f(\theta | \text{rest}) \) means the conditional distribution obtained by treating the parameters other than \( \theta \), together with the observed data, as fixed):

1. \( f(\beta_k | \text{rest}) \sim N_p(A_k a_k, A_k) \), \( a_k = \sum I(C_i = k)\sigma^{-2}_\beta \Lambda'_{i\gamma}\gamma_i^t + \Psi_0 \beta^* \) and \( A_k = ((\sum I(C_i = k)\sigma^{-2}_\beta \Lambda'_{i\gamma}\gamma_i), -1 + \Psi_0^{-1}) \), where \( I(\cdot) \) is the function equal to 1 if \( (\cdot) \) is true and 0 otherwise, and \( y^s_i = y_i - \Lambda_{i\gamma} \beta_k \).

2. \( f(\sigma^2_k | \text{rest}) \sim \text{INV} - \chi^2(n_i + 2, (n_i + 2a)^{-1}(\hat{s}^2_i + 2b)) \), where \( \hat{s}^2_i = \sum I(y_{it} - \Lambda_{i\gamma} [\beta_k + b_j])^2 \).

3. \( f(b_j | \text{rest}) \sim N_p \left( (\Lambda_{i\gamma} \beta_k + \sigma^{-2}_\beta \psi^{-1})^{-1} A'_{ij}\hat{y}_{ij}, (\Lambda_{i\gamma} \beta_k + \sigma^{-2}_\beta \psi^{-1})^{-1} \right) \), where \( \hat{y}_{ij} = y_{ij} - \Lambda_{i\gamma} \beta_k \).

4. \( f(\psi | \text{rest}) \sim \text{INV-WISHART}(N + v_1, S_0) \) where \( S_0 = \sum a_j b_j \).

5. Draw \( \delta_k \) via the SIR algorithm, where the mean and variance of the dominating distribution are obtained via the Newton–Raphson algorithm and \( p(\delta_k | \text{rest}) \propto e^{-\frac{1}{2} x \Sigma^{-1} d} \prod_{i=1}^{n} \left( e^{A_{i\gamma}|\gamma_i + d_i|_{\nu_i}}/(1 + e^{A_{i\gamma}|\gamma_i + d_i|_{\nu_i}}) \right) ; \delta_k \) is fixed at zero for identifiability.

6. Draw \( d_i \) via the SIR algorithm, where the mean and variance of the dominating distribution are obtained via the Newton–Raphson algorithm and \( p(d_i | \text{rest}) \propto e^{-\frac{1}{2} x \Sigma^{-1} d} \prod_{i=1}^{n} \left( e^{A_{i\gamma}|\gamma_i + d_i|_{\nu_i}}/(1 + e^{A_{i\gamma}|\gamma_i + d_i|_{\nu_i}}) \right) . \)

7. \( f(\Sigma | \text{rest}) \sim \text{INV-WISHART}(N + v_2, S_\Sigma) \) where \( S_\Sigma = \sum d_i d_i^t + S_2 \).

8. \( f(\gamma_k | \text{rest}) \sim N_0(A_k a_k, A_k) \), where \( a_k = ((\hat{V}_{\gamma_k})^{-1}\gamma_\kappa + \Sigma_0^{-1}\gamma*) \) and \( A_k = ((\hat{V}_{\gamma_k})^{-1} + \Sigma_0^{-1})^{-1} \).

Here \( \gamma_\kappa \) is obtained by logistic regression of \( v \) on \( \Lambda_k \) using \( d_i \) as an offset, and \( \hat{V}_{\gamma_k} = \text{Cov}(\gamma_k) \) is given by the inverse of the negative of the associated Hessian matrix.

9. \( f(C_i | \text{rest}) \sim \text{MULTI}(1, \pi^t_i) \), where \( \pi^t_i = (\hat{\pi}_i1 / \sum \hat{\pi}_ik, \ldots, \hat{\pi}_K / \sum \hat{\pi}_ik) \), and

\[
\hat{\pi}_ik = \frac{e^{-(y_{it} - \Lambda_{i\gamma} \beta_k)} |\beta_k + b_j|^2 / 2\sigma^2_k}{\int e^{-(y_{it} - \Lambda_{i\gamma} \beta_k)} |\beta_k + b_j|^2 / 2\sigma^2_k} \times \left( \frac{\delta_i x_i}{1 + e^{\Lambda_{i\gamma} |\gamma_i + d_i|_{\nu_i}}} \right)^{1-\nu_0} \times \left( \frac{\sum \delta_i x_i}{\sum \delta_i x_i} \right)
\]

where \( \delta_k = 0 \).

Obtaining initializing estimates

Ignoring the subject-level clustering and regressing the positive affect measures \( y \) on the time-of-observation matrix \( \Lambda \), yields within- and between-subject residual mean square error estimates that provide initial starting values of \( \sigma^2 \) and \( \Psi \). Initial estimates of \( b_i \) are then simply drawn from an \( N_m(0, \Psi) \) distribution. Initial class assignments \( C_i \) are drawn from a discrete uniform \([1, K]\) distribution. Conditional on \( C_i \), then, initial estimates of \( \gamma_k \) are made via logistic regression of \( v \), the negative event indicators, on time-of-observation matrix \( \Lambda_k \). The initial draw of \( d_i \) was made from an \( N_l(0, I) \) distribution.
Bayesian model to identify trajectories of positive affect and negative events

REFERENCES


Bayesian model to identify trajectories of positive affect and negative events


[Received August 27 2003; first revision February 23, 2004; second revision June 10, 2004; third revision August 6, 2004; accepted for publication August 16, 2004]