University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2006 *Paper* 11

Nested Markov Compliance Class Model in the Presence of Time-Varying Noncompliance

Julia Y. Lin* Thomas R. TenHave† Michael R. Elliott‡

Nested Markov Compliance Class Model in the Presence of Time-Varying Noncompliance

Julia Y. Lin, Thomas R. TenHave, and Michael R. Elliott

Abstract

We consider a Markov structure for partially unobserved time-varying compliance classes in the Imbens-Rubin (1997) compliance model framework. The context is a longitudinal randomized intervention study where subjects are randomized once at baseline, outcomes and patient adherence are measured at multiple follow-ups, and patient adherence to their randomized treatment could vary over time. We propose a nested latent compliance class model where we use time-invariant subject-specific compliance principal strata to summarize longtudinal trends of subject-specific time-varying compliance patterns. The principal strata are formed using Markov models that related current compliance behavior to compliance history. Treatment effects are estimated as intent-to -treat effects within the compliance principal strata.

Nested Markov Compliance Class Model in the Presence of Time-Varying Noncompliance

JULIA Y. LIN*

Department of Biostatistics and Epidemiology
University of Pennsylvania School of Medicine,
423 Guardian Drive, Philadelphia, Pennsylvania 19104, USA
email: jlin@cceb.med.upenn.edu, phone: 215.573.8950, fax: 215.573.1203

THOMAS R. TEN HAVE

Department of Biostatistics and Epidemiology
University of Pennsylvania School of Medicine, USA

MICHAEL R. ELLIOTT

Department of Biostatistics, University of Michigan, USA
Institute for Social Research, University of Michigan, USA

SUMMARY

We consider a Markov structure for partially unobserved time-varying compliance classes in the Imbens-Rubin (1997) compliance model framework. The context is a longitudinal randomized intervention study where subjects are randomized once at baseline, outcomes and patient adherence are measured at multiple follow-ups, and patient adherence to their randomized treatment could vary over time. We propose a nested latent compliance class model where we use time-invariant subject-specific compliance principal strata to summarize longitudinal trends of subject-specific time-varying compliance patterns. The principal strata are formed using Markov models that related current compliance behavior to compliance history. Treatment effects are estimated as intent-to-treat effects within the compliance principal strata.

KEY WORDS: Longitudinal compliance class model; Principal stratification; Markov; Latent transition model; Noncompliance; Geriatric depression.

*To whom correspondence should be addressed.

1 Introduction

In randomized intervention studies where interventions are administered repeatedly, subject adherence to the randomized treatment may vary over time. We consider a longitudinal compliance class model that uses nested principal stratification structure to characterize longitudinal compliance patterns over time within which intent-to-treat effects are estimated. We consider a Markov structure for the time-varying subject compliance behavior. We illustrate the model with analysis of the "Prevention of Suicide in Primary Care Elderly: Collaborative Trial" (PROSPECT) study (Bruce *et al.* 2004).

The PROSPECT study was a randomized intervention study targeted at elderly patients in primary care clinics with depression. There were two treatment groups: usual care and the intervention. In the usual care group, patients received standard care. In the intervention group, patients were assigned to meet with health specialists who educated patients, their families, and physicians about depression, treatment, and monitored adherence to treatment. Primary care clinics were randomized to the treatments rather than individual patients to prevent contamination of treatments between patients within the same clinic and for practicality. Patients were followed for two years from the initial randomization. Clinical depression outcome and adherence to randomized treatment were measured at 4, 8, 12, 18, and 24 months. There were 598 patients in the study. The clinical outcome of interest is the severity of depression measured by the Hamilton Depression Score (HAMD). The mean HAMD score at baseline is 18.1 with a standard deviation of 6.0. We consider an all-or-none treatment adherence measured by whether patients met with the health specialists at least once since the previous follow-up period. We are interested in investigating the effect of the intervention on depression severity controlling for treatment non-adherence.

In randomized intervention studies where subjects do not always adhere to the treatment to which they are randomized, subject noncompliance could confound the relationship between the treatment and the outcome. Therefore, it is important to account for subject noncompliance when estimating the effect of the treatment using principal stratification strategies (Frangakis and Rubin, 1999, 2002). Angrist *et al.* (1996) and Imbens and Rubin (1997) proposed to use compliance classes to describe subject compliance behaviors within which intent-to-treat (ITT) contrasts are made to estimate the causal effect of the treatment on the outcome.

In cross-sectional studies with two treatment arms, experimental treatment and control treatment, there are four possible compliance classes: compliers, always-takers, never-takers, and defiers. Compliers are those that would adhere to the treatment to which they are assigned; always-takers are those that would seek to receive the experimental treatment regardless of their treatment assignment; never-takers are those that would opt to receive the control treatment regardless of their treatment assignment; and defiers are those that would refuse the treatment to which they are assigned and choose to receive the other treatment. In this study design, subject compliance classes are latent. An individual assigned to the experimental treatment and receives the experimental treatment could be a complier or an always-taker; an individual assigned to the experimental treatment but receives the control treatment could be a never-taker or a defier; an individual assigned to the control treatment and receives the control treatment could be a complier or a never-taker; and an individual assigned to the control treatment but receives the experimental treatment could be an always-taker or a defier.

In studies, such as PROSPECT, where those assigned to the control treatment have no access to the experimental treatment, there are only compliers and never-takers. Always-takers do not exist in such study designs because those randomized

to the control treatment cannot receive the experimental treatment, and defiers do not exist for the same reason. The compliance classes for those assigned to the experimental treatment in this study design are observed. Subjects assigned to and receive the experimental treatment are compliers; subjects assigned to the experimental treatment but choose to receive the control treatment are never-takers. The compliance classes for those assigned to the control treatment are unobserved.

In this paper we propose an extension of the cross-sectional model in Imbens and Rubin (1997) to longitudinal settings. Yau and Little (2001) extended the Imbens and Rubin (1997) model to a longitudinal randomized intervention study where unemployed subjects received preventive intervention and their employment and mental health outcomes were measured. Although outcomes were measured repeatedly over time, the intervention was only administered once in the beginning of the study. Therefore, adherence to intervention was only recorded once and did not vary over time. Our proposed model allows treatment adherence to vary over time.

In longitudinal randomized intervention studies where treatments are applied repeatedly, subject treatment adherence could vary over time. In Frangakis *et al.* (2004), the randomized distance between the needle exchange truck and subjects' residence could change over time, and subject needle exchange behavior could also vary over time. HIV (human immunodeficiency virus) seroconversion status were recorded at multiple follow-ups. The goal of the analysis was to determine the effect of exchanging needles on HIV seroconversion, taking into account the history of compliance behaviors. This model differs from our proposed model in two ways: 1) we do not allow randomization status to change over time; 2) we propose a nested model structure that uses subject-specific time-invariant principal strata to summarize subject-specific time-varying compliance behavior.

In the presence of time-varying compliance behaviors, it may be useful to consider longitudinal compliance behavior patterns when examining longitudinal outcomes. Subjects with different compliance trajectories may differ in treatment outcomes. We may make inferences on different longitudinal compliance patterns and the longitudinal outcomes associated with those patterns. In a study like PROSPECT where there are two possible compliance classes and 5 follow-up visits, we have 32 (2^5) possible compliance patterns. It may be impractical and not clinically meaningful to look at the longitudinal outcomes in all of the 32 patterns. Hence, it may be more helpful to have summary measures of the longitudinal compliance patterns in the data, and look at longitudinal outcomes within broader latent classes.

This paper extends a nested latent class model proposed by Lin *et al.* (in preparation) to accommodate time-varying latent compliance classes by specifying broader principal strata that summarize the compliance classes. The nested latent class model involves two levels of compliance class models. The first level uses subject-specific time-varying compliance classes to describe the time-varying treatment adherence; the second level uses subject-specific time-invariant compliance "superclasses" to summarize the longitudinal patterns of compliance classes. The ITT effect of the intervention stratified on compliance superclass, or principal effect (Frangakis and Rubin 2002), is estimated to control for longitudinal subject treatment noncompliance. The model assumes that compliance classes at each time point within an individual are independent from each other conditional on the individual's compliance superclass and baseline covariates. In other words, knowing the compliance superclass and subject baseline characteristics, the history of compliance behaviors does not provide any more information on the compliance behaviors. This may be a strong assumption which we now propose to assess with a Markov model for the time-varying compliance classes.

Collins and Wugalter (1992) first proposed latent transitional models to model dynamic latent classes where subjects advance through a sequence of latent stages and to look at the relationship between the latent stages. The model estimates the probabilities of remaining in the same stage or advancing to the next stage conditional on previous stage. This method was applied to the process of math learning (Collins and Wugalter, 1992) and stages of smoking cessation (Velicer *et al.*, 1996). Extensions by Reboussin *et al.* (1999), Humphreys and Janson (2000), and Reboussin *et al.* (2002) incorporated covariates to estimate the latent transitional probabilities using logistic regression. This latent transitional model was also extended to accommodate multiple indicators for estimating latent stages (Reboussin *et al.*, 1998) and mixed outcomes (Miglioretti, 2003).

Markov transition models have been used to study cost effectiveness of depression treatments in treating patients with major depression (Nuijten *et al.*, 1995; Revicki *et al.*, 1995). In behavioral health studies where treatments may be dynamic, it may be important to consider previous treatment adherence to better characterize course of treatment. Previous adherence to treatments may provide insights on how patients would respond to treatments in the future.

In this paper, we extend the longitudinal nested compliance class model in Lin *et al.* (in preparation) and relax the conditional independence assumption of the compliance classes given compliance superclass and baseline covariates. We assume a first-order Markov structure for the compliance classes given superclass and baseline covariates where compliance behaviors are assumed to depend on the compliance class in the previous time point. Modelling the Markov structure of the time-varying compliance classes will allow us to: 1) utilize information from history of compliance to predict compliance behaviors; 2) examine how history of compliance relate to compliance behavior.

We will define notation, discuss assumptions, principal effect, the parametric model, parameter estimation, the handling of missing outcomes, and assessment of model fit in Section 2. Then we will proceed to discuss the analysis results in Section 3, and finally make concluding remarks in Section 4.

2 Nested Compliance Class Model

2.1 Notation

Let Z_i denote the randomization status for subject i where $i = (1, \cdots, N)$, and $Z_i \in (0, 1)$ for randomized to the usual care and the intervention group, respectively. Similarly, let D_{ij} denote the time-varying treatment received for subject i at time j where $j = (1, 2, 3, 4, 5)$ for 4, 8, 12, 18, and 24 months, respectively, and $D_{ij} \in (0, 1)$ for usual care and intervention, respectively. Note that Z_i does not have the subscript j because we are restricting to designs where randomization does not change over time. Let Y_{ij} denote the observed outcome for subject i at time j. We use $\mathbf{Z}$, $\mathbf{D}$, $\mathbf{Y}$ to denote vectors of Z_i, D_{ij}, and Y_{ij}.

Following Little and Rubin (2000), we use $Y_{ij}(Z)$ to denote the partially latent potential outcome, outcome that would have been observed, for subject i at time j if randomized to treatment Z. Let C_{ij} denote membership of the partially latent compliance classes for subject i at time j. In the PROSPECT study, since those randomized to the usual care group have no access to the intervention, there are only two possible compliance classes: compliers and never-takers; therefore, $C_{ij} = (c, n)$. We use $\mathbf{C}$ to denote the vector of C_{ij}.

The proposed principal stratification strategy uses compliance "superclasses" to summarize the longitudinal compliance patterns in the data within which we can stratify on and compare potential outcomes. It precludes the confounding when

stratifying on observed post-randomization compliance patterns. Let U_i denote membership of the latent superclass for subject i, where $U_i = (1, \cdots, K)$ for assumed K numbers of latent superclasses. We use $\mathbf{U}$ to denote the vector of U_i.

Subject-level baseline covariates $\mathbf{A}_i$ and $\mathbf{Q}_i$ are used in modelling the outcome and compliance probabilities, respectively. We use $\mathbf{A}$ and $\mathbf{Q}$ to denote the vector of $\mathbf{A}_i$ and $\mathbf{Q}_i$.

We use upper case letter to denote random variables or indices of potential outcomes (e.g. $Y_{ij}(Z)$), and lower case letter to denote realized or observed values of the random variables or indices (e.g. $Z_i = z$).

2.2 Assumptions

We make the randomization (Rubin, 1978), stable unit-treatment value (SUTVA; Rubin, 1986), and model assumptions to identify causal model parameters. We assume that potential outcomes, latent compliance classes, and latent compliance superclasses (which are assumed to be baseline characteristics) are independent of randomization assignment status conditional on baseline covariates. We make the no interference assumption of the SUTVA and assume that the potential outcome of an individual is not influenced by the treatment assignment of another individual. We also make the consistency assumption of the SUTVA which assumes that the potential outcome of a certain treatment will be the same regardless of the treatment assignment mechanism. It implies that the observed outcome is a function of the potential outcomes and treatment assignment: $Y_{ij} = Z_i * Y_{ij}(1) + (1 - Z_i) * Y_{ij}(0)$. The SUTVA assumption is violated when there are interference between subjects or when there are versions of treatments not represented by the treatment indicator variable.

2.3 Principal Effects

We utilize compliance superclasses to summarize the longitudinal compliance patterns, and estimate ITT effects stratified on these superclasses. A compliance superclass is a latent subject-level principal stratum that is time-invariant, and is considered to be a pre-randomization characteristic which would allow us to model potential outcomes.

Our effect of interest is the principal effect of treatment assignment on the outcome within a compliance superclass at time j:

$$E[Y_{ij}(1)|U_i = k] - E[Y_{ij}(0)|U_i = k] \tag{1}$$

It is an ITT contrast stratified on the compliance superclass. It allows us to consider the effect of treatment randomization controlling for longitudinal compliance behavior. Since the superclasses defined here create baseline principal strata summarizing longitudinal compliance behaviors and do not represent specific longitudinal compliance patterns, the ITT contrasts sacrifice straightforward causal interpretations.

The principal effect can be defined by observed outcomes under the randomization and the SUTVA consistency assumption:

$$\begin{aligned} & E[Y_{ij}(Z=1)|U_i = k] - E[Y_{ij}(Z=0)|U_i = k] \\ = \;& E[Y_{ij}(Z=1)|Z_i = 1, U_i = k] - E[Y_{ij}(Z=0)|Z_i = 0, U_i = k] \\ = \;& E[Y_{ij}|Z_i = 1, U_i = k] - E[Y_{ij}|Z_i = 0, U_i = k] \end{aligned} \tag{2}$$

The first equal sign follows from the randomization assumption, which says that randomization is independent of baseline characteristics (e.g. potential outcomes)

conditional on baseline covariates (e.g. compliance superclass). The second equal sign follows from the SUTVA consistency assumption which implies that the observed outcome given treatment assignment z is the potential outcome for treatment assignment $Z = z$.

2.4 Parametric Model

Lin *et al.* (in preparation) proposed a conditional independence (CI) model where longitudinal compliance classes within an individual were assumed to be independent given compliance superclass and baseline covariates. Under the current proposed method we relax the CI assumption. We assume compliance classes are dependent on the compliance classes at one or more previous time points, the compliance superclass, and baseline covariates.

Following Lin *et al.* (in preparation), we assume outcomes within individuals are independent given randomization, time-varying compliance class, baseline covariates, and subject-level random effect.

$$\begin{aligned} \left(Y_{ij}|C_{i1},\cdots,C_{ij},Z_i=z,\mathbf{A}_i,\mathbf{W}_i,\boldsymbol{\lambda},\zeta(t,j),\boldsymbol{\gamma},\boldsymbol{\varphi}_i,\sigma^2\right) &\overset{ind}{\sim} N(\mu_{ijz},\sigma^2) \\ \mu_{ijz} &= \sum_{t=1}^{j}\left[\sum_{\eta'} I(C_{it}=\eta',Z_i=z)\lambda_{t\eta' z}\zeta(t,j)\right] + \mathbf{A}_i^T\boldsymbol{\gamma} + \mathbf{W}_i^T\boldsymbol{\varphi}_i \end{aligned} \tag{3}$$

where $\boldsymbol{\lambda}$ denotes the vector of $\lambda_{t\eta' z}$ parameters for $t \leq j$ that describe the compliance-class specific ITT effect of the treatment on the outcome, and $\zeta(t,j)$ modifies that ITT effect at time t on the outcome at time j that we will discuss more later. The conditional mean of the outcome has three components: compliance class-specific effect of randomization, the effect of baseline covariates, and the subject-specific random effects to account for within-subject correlation in the outcomes. The compliance class-specific effect of randomization on out-

come is represented by $\sum_{t=1}^{j}\left[\sum_{\eta'} I(C_{it}=\eta', Z_i=z)\lambda_{t\eta' z}\zeta(t,j)\right]$. The effect of the baseline covariates on the outcome is represented by $\mathbf{A}_i^T\boldsymbol{\gamma}$ where $\mathbf{A}_i$ denotes the vector of baseline covariates of subject i, and the column vector $\boldsymbol{\gamma}$ denotes the corresponding coefficients. The random effects $\boldsymbol{\varphi}_i$ is used to account for within-subject correlation in the outcomes, where $\mathbf{W}_i$ denotes the random effect design matrix for subject i. Previous analysis indicated that the within-clinic correlation was small (0.075), hence was ignored in this analysis. In this paper, we consider a random subject-level intercept model.

In some studies, it may be reasonable to consider decay of treatment effects over time. For example, the concentration of drugs in the body diminishes over time and may not completely dissipate before the next administration. Therefore, it would be important to model the decay of effect of the previous treatment when estimating the effect of the treatment. In the PROSPECT study we may consider the decay of the effect of previous compliance behavior on the outcome. It is conceivable that information ascertained in meetings with health specialists may have lasting effects on the subjects and their treatment outcomes. We can use the parameter $\zeta(t,j)$ to modify the relationship between compliance behavior at time t on the outcome at time j. We can assume a transient relationship where the outcome is only dependent on the current compliance behavior (i.e. $\zeta(t,j)=I(t=j)$); assume a non-transient relationship where ITT effect is dependent on current and all prior compliance behavior (i.e. $\zeta(t,j)=I(t\leq j)$); or assume a decaying relationship where the outcome is dependent on current and all prior compliance behavior, but the effect diminishes as time lag increases (i.e. $\zeta(t,j)=e^{-\tau(j-t)}$ where $\tau>0$). Preliminary analysis of the data using a decay model suggested $\tau\to\infty$, or a transient relationship between compliance behavior and ITT effect.

Hence, we consider the transient model:

$$\mu_{ijz} = \sum_{\eta'} [I(C_{ij} = \eta', Z_i = z)\lambda_{j\eta' z}] + \mathbf{A}_i^T \boldsymbol{\gamma} + \mathbf{W}_i^T \boldsymbol{\varphi}_i \tag{4}$$

Departing from Lin *et al.* (in preparation), we propose a Markov compliance class (MCC) model for the time-varying compliance classes, where the compliance classes are dependent on past compliance behavior. Following Lin *et al.* (in preparation), we assume that compliance superclass is an underlying factor that drives subject compliance over time. We model compliance class at the first time point conditional on compliance superclass and baseline covariates $\mathbf{Q}_i$ using logit models: $P(C_{i1} = \eta | U_i = k, \mathbf{Q}_i) = \omega_{k\eta}(\mathbf{Q}_i)$ and $\omega_{k\eta}(\mathbf{Q}_i) = exp(\alpha_{0k\eta} + \boldsymbol{\alpha}_{1\eta}\mathbf{Q}_i)/[\sum_{\eta'} exp(\alpha_{0k\eta'} + \boldsymbol{\alpha}_{1\eta'}\mathbf{Q}_i)]$ where $\sum_\eta \omega_{k\eta}(\mathbf{Q}_i) = 1 \; \forall k$. We constrain $\alpha_{0k\eta}$ and $\boldsymbol{\alpha}_{1\eta}$ for one of the compliance class η to be 0 for identifiability.

We assume subject compliance superclass $(U_i = k) \sim Multinomial(1, p_k)$, where $\sum_k p_k = 1$. Compliance superclass between subjects are assumed to be independent: $f(\mathbf{U}) = \prod_{i=1}^N f(U_i = k)$ for $k = 1, \cdots, K$ where $f(.)$ denotes the distribution function.

We utilize latent transitional models (Collins and Wugalter, 1992) to model the Markov process of compliance classes across time. In this paper we consider a non-stationary first-order Markov compliance model. The number of model parameters in multiple-order Markov models increase exponentially without additional constraints such as stationarity. Because of the lack of good predictors of compliance transitions, we assume that there are no associated covariates. Covariates can be incorporated using logit models as in Reboussin *et al.* (1999) and Humphreys and Janson (2000). We assume the compliance class transitions $(C_{ij} = \eta | C_{i,j-1} = \eta', U_i = k) \sim Multinomial(1, \pi_{kj\eta'\eta})$, where $\sum_\eta \pi_{kj\eta'\eta} = 1 \; \forall k, j, \eta'$. The joint distribution of the compliance classes given

compliance superclass then becomes:

$$P(C_{i1}, \cdots, C_{i5}|U_i, \mathbf{Q}_i) = P(C_{i1}|U_i, \mathbf{Q}_i)P(C_{i2}|C_{i1}, U_i) \cdots P(C_{i5}|C_{i4}, U_i) \quad (5)$$

If the compliance class and compliance superclass memberships, and the missing outcomes are known, the joint distribution of the complete data for subject i given the model specifications is as follows:

$$\begin{aligned} & f(Y_{i1}, \cdots, Y_{i5}, \boldsymbol{\varphi}_i, C_{i1}, \cdots, C_{i5}, U_i|Z_i, \mathbf{A}_i, \mathbf{Q}_i, \mathbf{W}_i, \boldsymbol{\theta}) \\ &= f(Y_{i1}, \cdots, Y_{i5}|\boldsymbol{\varphi}_i, C_{i1}, \cdots, C_{i5}, U_i, Z_i, \mathbf{A}_i, \mathbf{Q}_i, \mathbf{W}_i, \boldsymbol{\theta}) \times \\ & \quad f(\boldsymbol{\varphi}_i|C_{i1}, \cdots, C_{i5}, U_i, Z_i, \mathbf{A}_i, \mathbf{Q}_i, \mathbf{W}_i, \boldsymbol{\theta}) \times \\ & \quad f(C_{i1}, \cdots, C_{i5}|U_i, Z_i, \mathbf{A}_i, \mathbf{Q}_i, \mathbf{W}_i, \boldsymbol{\theta}) f(U_i|Z_i, \mathbf{A}_i, \mathbf{Q}_i, \mathbf{W}_i, \boldsymbol{\theta}) \quad (6) \\ &= f(Y_{i1}, \cdots, Y_{i5}|C_{i1}, \cdots, C_{i5}, Z_i, \mathbf{A}_i, \mathbf{W}_i, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\varphi}_i, \sigma^2) \times \\ & \quad f(\boldsymbol{\varphi}_i|\Sigma_\varphi) f(C_{i1}, \cdots, C_{i5}|U_i, \mathbf{Q}_i) f(U_i) \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2, \Sigma_\varphi)$

2.5 Estimation

We use a Bayesian Markov Chain Monte Carlo (MCMC) method to estimate model parameters. Let $\beta = [\lambda_{1c0}, \cdots, \lambda_{5n1}, \boldsymbol{\gamma}]^T$ denote the fixed effects. We assume the conjugate priors $\beta \sim MVN(\mu_\beta, \Sigma_\beta)$ and $\sigma^2 \sim Inv-\chi^2(df = \nu_\sigma, \psi)$. We assume $\boldsymbol{\varphi}_i \sim MVN(\mathbf{0}, \Sigma_\varphi)$ for the subject-level random effects, and the hyper-prior $\Sigma_\varphi \sim Inv-Wishart(df = \nu_\varphi, \Gamma)$. For compliance superclass and compliance class probabilities, we assume the priors $(p_1, \cdots, p_K) \sim Dirichlet(a_1, \cdots, a_K)$, $\boldsymbol{\alpha} \sim MVN(0, \Sigma_\alpha)$, and $(\pi_{kj\eta' c}, \pi_{kj\eta' n}) \sim Dirichlet(b_c, b_n)$ $\forall k, j, \eta'$. Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990; Imbens and Rubin, 1997; Gelman *et al.*, 2004) is used to otbain draws from the posterior distributions of the parameters. The posterior distributions from which the model parameters

are drawn are presented in the Appendix. The posterior distribution $(\boldsymbol{\alpha}|\mathbf{C},\mathbf{U},\mathbf{Q})$ where $\boldsymbol{\alpha}=[\alpha_{01c},\cdots,\alpha_{0Kn}\boldsymbol{\alpha}_{1c},\boldsymbol{\alpha}_{1n}]$ is not of a known parametric form. Therefore, we use the Metropolis-Hasting algorithm (Hastings, 1970, Gelman *et al.*, 2004) to draw the $\boldsymbol{\alpha}$ parameters.

2.6 Missing Outcome Imputation

There are missing outcomes (Y_{ij}) present in the PROSPECT data. We assume a latent ignorable missing data mechanism (LIMD; Peng *et al.*, 2004) for the missing outcomes, which assumes missing at random given latent compliance class and covariates. At each iteration of the MCMC procedure, we impute the missing outcomes conditional on compliance classes, treatment randomization, baseline covariates, and subject-level random effects. We draw missing outcome Y_{ij}^{mis} for subject i at time j from its predictive distribution given current values of parameters C_{ij}, $\lambda_{j\eta z}$, $\boldsymbol{\gamma}$, $\boldsymbol{\varphi}_i$, σ^2, and vector of observed outcomes $\mathbf{Y}^{obs}$.

$$\begin{aligned}(Y_{ij}^{mis}|\mathbf{Y}^{obs},C_{ij},Z_i=z,\mathbf{A}_i,\mathbf{W}_i,\lambda_{j\eta z},\boldsymbol{\gamma},\boldsymbol{\varphi}_i,\sigma^2)&\sim N(\mu^*_{ijz},\sigma^2)\\ \mu^*_{ijz}&=\sum_{\eta'}[I(C_{ij}=\eta',Z_i=z)\lambda_{j\eta' z}]+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i\end{aligned} \tag{7}$$

2.7 Model Fit Assessment

We compare the fits of the MCC model to the CI model in Lin *et al.* (in preparation) by comparing the posterior predictive distributions (PPD; Gelman *et al.*, 2004) of the time-varying compliance classes between the MCC and the CI models. Let G_m denote the number of individuals in the m^{th} of the 32 possible longitudinal compliance patterns and let κ_m be the estimated probability of exhibiting the m^{th}

longitudinal compliance pattern. We consider the χ^2-type statistics:

$$S^{obs} = \sum_m \frac{(G_m^{obs} - N\kappa_m)^2}{N\kappa_m(1-\kappa_m)} \text{ and } S^{rep} = \sum_m \frac{(G_m^{rep} - N\kappa_m)^2}{N\kappa_m(1-\kappa_m)} \tag{8}$$

where G_m^{obs} is the observed statistics and G_m^{rep} is the repeated statistic obtained from draws of the parameters generated by the Gibbs sampler. The PPD p-value is then given by:

$$\frac{\sum_l I[(S^{obs})^l < (S^{rep})^l]}{\sum_l 1} \tag{9}$$

where $(S^{obs})^l$ and $(S^{rep})^l$ denote the S^{obs} and S^{rep} from l^{th} Gibbs draw. A PPD p-value close to 0.50 indicates good fit of the model to the data.

3 Results

We demonstrate the MCC model with the PROSPECT data and compare the results to the CI model assuming three superclasses. There are unrecorded treatment received (D_{ij}) in the data, of which we assume are 0, or have not met with health specialists. Imputation of missing outcomes is as described in Section 2.6. In the PROSPECT study, those randomized to the usual care group do not have access to the intervention; therefore, there are only two compliance classes: compliers and never-takers. Goodman (1974) suggests that we can only identify at most 3 latent compliance superclass given 5 dichotomous compliance classes. Lin *et al.* (in preparation) showed that the 3-superclass CI model fits the data better than the 2-superclass CI Model. Our effect of interest is the effect of the health specialists on the severity of depression. More specifically, we are interested in the principal ITT effect of the intervention on the outcome stratified on compliance

superclass.

$$\begin{aligned} & E[Y_{ij}(Z=1)|U_i=k] - E[Y_{ij}(Z=0)|U_i=k] \\ = & \sum_{\eta'} (\lambda_{j\eta'1} - \lambda_{j\eta'0})\, P(C_{ij}=\eta'|U_i=k) \end{aligned} \tag{10}$$

In this analysis we let $\mathbf{A}_i$ be the baseline HAMD score and baseline suicidal ideation. We want to control for the baseline HAMD because we are interested in the change in HAMD scores from baseline. Treatment randomization failed to balance the proportion of subjects with baseline suicidal ideation in the treatment groups; therefore, we want to control for it in modelling the outcome. We let $\mathbf{Q}_i$ be the baseline HAMD score in estimating the compliance probabilities in the CI model and in estimating the initial compliance probabilities in the MCC model.

We use noninformative priors in the Bayesian MCMC estimation of the model parameters since we do not have strong prior inclinations. Following Garrett and Zeger (2000) and Ten Have *et al.* (2004) we assume $\boldsymbol{\alpha} \sim MVN(\mathbf{0}, \Sigma_\alpha = diag(50, 4))$. The difference in variance component in the priors reflect the different scaling of the covariates. A larger variance is used for binary covariates (i.e. intercept) and a smaller variance is used for continuous covariates (i.e. baseline HAMD score; Garrett and Zeger, 2000, Ten Have *et al.*, 2004). The identifiability of the $\boldsymbol{\alpha}$ parameters are checked by comparing the prior and the posterior distributions (Garrett and Zeger, 2000). We assume the prior $(\pi_{kj\eta'c}, \pi_{kj\eta'n}) \sim Dirichlet(0.01, 0.01) \forall k, j, \eta'$ for the transitional probabilities. This is equivalent to adding 0.01 subject to each of the $(C_{i,j-1} = \eta', C_{ij} = \eta | U_i = k)$ groups. We assume $\boldsymbol{\beta} \sim MVN(\mu_\beta = \mathbf{0}, \Sigma_\beta = 1000 \times \mathbf{I})$ and $\sigma^2 \sim Inv - \chi^2(\nu_\sigma = 1, \psi = 1/10)$. For the random effect variance parameter we assume $\Sigma_\varphi \sim Inv - \chi^2(\nu_\varphi = 1, \Gamma = 1/10)$. We assume the prior $(p_1, \cdots, p_K) \sim Dirichlet(1, \cdots, 1)$, assigning a priori 1 subject to each of the K superclasses.

We used the Gelman-Rubin $\hat{R}$ statistic (Gelman *et al.*, 2004, pp.296-297) to assess the convergence of the MCMC chains and $\hat{R} < 1.1$ is accepted as evidence of convergence. We ran 3 chains of the CI model for 10,000 iterations each and the first 1,000 iterations was discarded as burn-in, and we ran 3 chains of the MCC model for 150,000 iterations each and the first 75,000 iterations was discarded as burn-in. The maximum $\hat{R}$ was 1.05 and 1.08 for the CI and the MCC models, respectively.

We will present the results under the CI model as specified in Lin *et al.* (in preparation), then the results under the MCC model, follow by comparison of the two models. We can assess the conditional independence assumption made under the CI model by comparing the fit of the CI model to the fit of the MCC model to the data. We also compare the results under different model assumptions for the time-varying compliance classes.

3.1 Conditional Independence Model

For comparison, results under the CI model as described in Lin *et al.* (in preparation) are displayed in Tables 1 and 2. Table 1 shows the time- and superclass-varying compliance probabilities from the CI model assuming the average baseline HAMD of 18.1, and table 2 shows the ITT effect of randomization on the outcome within each compliance superclass controlling for the baseline HAMD and baseline suicidal ideation.

Table 1 shows that the first superclass under the CI model consists of subjects who are noncompliant at the 4-month follow=up and become even more noncompliant for the remainder of the study (low compliers). The second superclass consists of subjects who are highly compliant for the first 3 months and rapidly become noncompliant (decreasing compliers). The third superclass consists of subjects

who are highly compliant but become slightly less compliant at the last follow-up visit (high compliers). About 28% of the subjects in the PROSPECT study are low compliers with 95% credible interval (0.23,0.33), 16%(0.12,0.22) are decreasing compliers, and 56%(0.50,0.62) are high compliers.

[Table 1 about here.]

The posterior mean and the 95% credible interval of the log odds of compliance for every unit increase in baseline HAMD is 0.003(-0.04,0.05) suggesting those with more severe depression at baseline (higher baseline HAMD) are more likely to comply to treatment assignment than those with less severe depression at baseline.

The within-superclass ITT contrasts of (10) are shown Table 2. The contrasts suggest strong direct effect of randomization at the 4-month follow-up in the low complier superclass, which consists of largely never-takers, who are unlikely to meet with health specialists regardless of the treatment assigned. The direct effect of randomization on the outcome could be due to the unblinded nature of the intervention. The presence of the health specialists in the clinics may influence the behaviors of the health care providers and patients. The possible direct effect of randomization seems to dissipate over time. At the end of the first year, only the high compliers randomized to the intervention group, who are still highly likely to meet with a health specialist when assigned to the intervention, showed greater reduction in HAMD relative to high compliers in usual care. At the end of the 2-year study, none of the superclasses show strong ITT effects on depression.

[Table 2 about here.]

3.2 Markov Compliance Class Model

The Markov compliance class model relaxes the conditional independence assumption of the time-varying compliance class given compliance superclass and baseline covariates, and instead, assumes a first-order Markov structure for the time-varying compliance classes given compliance superclass. Again, we assume 3 compliance superclasses.

The posterior means of the log odds of compliance at 4 months and their associated 95% credible intervals within each of the three superclasses controlling for baseline HAMD under the MCC model are -0.52(-1.87,0.81), -3.61(-15.56,4.37), and 4.99(1.11,13.69) for the first, second, and third superclass, respectively. This suggests that at the 4-month follow-up, those in the first and second superclasses are less likely to comply with their treatment assignment, and those in the third superclass are more likely to comply with their treatment assignment. In our model we assume that the association between baseline HAMD and compliance probability at 4 month is the same across all three superclasses. The log odds of 4 month compliance for a unit increase in the baseline HAMD is 0.07(0.01,0.13) suggesting that those with more severe depression at baseline are more likely to comply with treatment assignment.

Table 3 shows the time-varying compliance probabilities when we assume an average baseline HAMD score of 18.1. The first superclass consists of subjects who are slightly more likely to comply with assigned treatment at the 4-month follow-up than not comply, and compliance decreases over time (increasing noncompliers). The second superclass consists of subjects who exhibit erratic compliance behavior with abrupt increases and decreases in compliance probabilities (erratic compliers). The third superclass consists of subjects who are highly compliant during the first 18 months of the study with a slight drop off in compliance during the

last 6 months (high compliers). The mean posterior probabilities and their associated 95% credible intervals of membership in the increasing noncomplier, erratic complier, and high complier superclasses are 0.42(0.25,0.56), 0.04(0.00,0.15), and 0.54(0.42,0.72), respectively.

[Table 3 about here.]

The latent transitional probabilities of the time-varying compliance within each superclass in Tables 4 shows that increasing noncompliers and high compliers are more likely to stay in the complier class if they are in the complier class in the previous time point than if they are in the never-taker class and switch to complier class. Subjects in the high complier superclass are more likely to transition to the complier class than subjects in the increasing noncomplier superclass. We do not see any clear patterns in the transitional probabilities of the erratic compliers.

[Table 4 about here.]

The posterior means and credible intervals of (10), the within-compliance superclass ITT contrasts, in Table 5 show an ITT effect at 4 months in the erratic compliers, suggesting a strong direct effect of randomization. The erratic complier superclass consists of mostly never-takers at 4 months, who are unlikely to meet with health specialists if assigned to the intervention. This direct effects seems to dissipate over time. We also see an ITT effect in the high compliers at 4 months, suggesting an effect of the intervention. The high complier superclass consists of almost entirely compliers at 4 months, who are likely to meet with health specialists if assigned to the intervention. Consistent with the results under the CI model, at the end of the first year we see a greater reduction in HAMD in the high compliers assigned to the intervention relative to the high compliers

assigned to the usual care. It suggests that meeting health specialists help improve depression. However none of the superclasses show strong ITT effects on depression.

[Table 5 about here.]

3.3 Model Comparison

In this section we compare the superclasses identified under the CI and the MCC compliance class structures. Under both compliance class structures we identified a superclass of high compliers, who are highly compliant throughout the study period with slight decrease in time-varying compliance at the last follow-up visit. We also identified a superclass that exhibits decreasing time-varying compliance; however, the compliance probability under the CI model starts out much higher at 4 months and decreases at a faster rate over subsequent visits than the compliance probability under the MCC model. Under the CI model we identified a superclass of subject who are noncompliant, with no clear time trend in their compliance probabilities. Under the MCC model we identified a superclass of subjects exhibiting erratic compliance behavior with fluctuating compliance probabilities and no clear trend in their compliance class transitions.

We saw similar within-compliance superclass ITT effects under both the CI and the MCC models. The ITT effects were larger in noncompliant subjects than compliant subjects at the 4-month follow-up suggesting a direct effect of randomization early on. This is most evident in the low complier superclass under the CI model and the erratic complier superclass under the MCC model, both of which consist of mostly never-takers at 4 months, who exhibit the largest ITT effects. However, this direct effect seems to dissipate over time. At the 24-month follow-

up we see the largest ITT effect in the high complier superclass under both the CI and the MCC models, which consist of mostly compliers. Although none of the ITT effects were strong in any of the superclasses under both models by the end of the study.

Assessment of the fit of the posterior predictive distribution to the data using the χ^2-type statistics in (8) suggests that the MCC model has a better fit than the CI model. The PPD p-value under the CI model is 0.0057 and 0.1549 under the MCC model.

4 Discussion

In Lin *et al.* (in preparation), a conditional independence model of the time-varying compliance classes was proposed that assumes the compliance classes within an individual are independent given compliance superclass and baseline covariates. People are creatures of habit. Those that exhibit history of compliance to an assigned treatment may be more likely to comply than those that exhibit history of noncompliance. In this paper, we proposed a Markov model of the time-varying compliance classes that assumes the compliance classes at each time point are dependent on the previous compliance behaviors, compliance superclass, and baseline covariates.

Under the MCC model we found that those who are more depressed at baseline are more likely to comply with their treatment assignment at 4 months. The same trend was also found under the CI model. Patients who are more depressed may be more eager to treat their depression and be more likely to adhere to their prescribed treatment. Physicians may also monitor patients with more severe depression more closely, making sure that the patients adhere to their treatments.

The proposed MCC model provides information on how treatment compliance relates to history of compliance that was not considered in the CI model. We found that most of the subjects in the PROSPECT study who complied to the treatment assignment in the previous follow-up period were more likely to comply again than those who were noncompliant during the previous visit.

When we compared the posterior predictive distributions under the MCC and the CI models to the data, we found that the MCC model fit the data much better than the CI model. The MCC model is a more flexible model; therefore, we expect it to have a better fit. Though the PPD p-value under the MCC models is still away from 0.50. In our future research, we plan to explore covariates related to compliance superclasses and time-varying compliance classes to improve the fit of the MCC model.

In our current analysis we model the relationship between baseline depression severity and compliance at 4 months, which indirectly models the relationship between baseline depression and longitudinal compliance. It is of clinical interest to identify patient characteristics that relate to treatment compliance. If clinicians can identify patients who are likely to comply to treatment over time and those less likely to comply, then clinicians may be able to target patients with particular attributes and tailor treatment for different patients to optimize patient treatment adherence and treatment outcomes. We are pursuing other predictors of longitudinal compliance in our current work.

One limitation of our proposed method is that the principal effect does not provide straightforward causal interpretation. Although the superclasses defined here provide convenient summaries of the longitudinal compliance patterns, they do not represent specific compliance patterns. Therefore, the ITT contrasts stratified on superclasses do not have straightforward causal interpretations.

5 Appendix: Conditional draws of the Gibbs sampler

Let $\mathbf{Y}_i$ and $\mathbf{C_i}$ denote the vectors of Y_{ij} and C_{ij} for subject i. For notational simplicity, let $X_{ij} = [I(C_{i1} = c, Z_i = 0), \cdots, I(C_{i5} = n, Z_i = 1), \mathbf{A}_i]$ denote the row vector of the fixed effect, and $\mathbf{X}_i$ denote the design matrix of the fixed effect for subject i with 5 (number of follow-ups) rows. Let $\boldsymbol{\beta} = [\lambda_{1c0}, \cdots, \lambda_{5n1}, \boldsymbol{\gamma}]$ denote the vector of coefficients corresponding to the fixed effect.

The distributions from which parameters are drawn at each iteration in the Gibbs sampling are as follows:

$$\begin{aligned}
\left(\boldsymbol{\beta}|\mathbf{X}, \mathbf{Y}, \mathbf{W}, \boldsymbol{\varphi}, \sigma^2, \boldsymbol{\mu}_\beta, \Sigma_\beta^{-1}\right) &\sim MVN(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) \\
\hat{\boldsymbol{\mu}} &= \frac{\sigma^{-2}\sum_{i=1}^N \mathbf{X}_i^T(\mathbf{Y}_i - \mathbf{W}_i^T\boldsymbol{\varphi}_i) + \Sigma_\beta^{-1}\boldsymbol{\mu}_\beta}{\sigma^{-2}\sum_{i=1}^N \mathbf{X}_i^T\mathbf{X}_i + \Sigma_\beta^{-1}} \\
\hat{\Sigma} &= (\sigma^{-2}\textstyle\sum_{i=1}^N \mathbf{X}_i^T\mathbf{X}_i + \Sigma_\beta^{-1})^{-1}
\end{aligned}$$

$$\begin{aligned}
(\sigma^2|\mathbf{X}, \mathbf{Y}, \mathbf{W}, \boldsymbol{\varphi}, \boldsymbol{\beta}, \nu_\sigma, \psi) &\sim Inv - \chi^2\left(df = 5N + \nu_\sigma, \tfrac{\sum_{i=1}^N F_i + \nu_\sigma\psi}{5N + \nu_\sigma}\right) \\
\text{where } F_i &= \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{W}_i^T\boldsymbol{\varphi}_i\right)^T\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{W}_i^T\boldsymbol{\varphi}_i\right)
\end{aligned}$$

$$\begin{aligned}
\left(\boldsymbol{\varphi}_i|\mathbf{X}_i, \mathbf{Y}_i, \mathbf{W}_i, \boldsymbol{\beta}, \sigma^2, \Sigma_\varphi, \right) &\sim MVN\left(\hat{\boldsymbol{\varphi}}_i\hat{V}_i, \hat{V}_i\right) \\
\hat{\boldsymbol{\varphi}}_i &= \frac{\mathbf{W}_i^T\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)}{\sigma^2} \\
\hat{\mathbf{V}}_i &= \left(\frac{\mathbf{W}_i^T\mathbf{W}_i}{\sigma^2} + \Sigma_\varphi^{-1}\right)^{-1}
\end{aligned}$$

$$(\Sigma_\varphi|\boldsymbol{\varphi}, \omega, \Gamma) \sim Inv - Wishart\left(df = \nu_\varphi + N, \sum_{i=1}^N \boldsymbol{\varphi}_i^T\boldsymbol{\varphi}_i + \Gamma\right)$$

$$(p_1, \cdots, p_K|\mathbf{U}, a_1, \cdots, a_k) \sim Dirichlet(r_1, \cdots, r_K)$$
$$\begin{aligned} r_1 &= \sum_{i=1}^{N} I(U_i = 1) + a_1 \\ r_K &= \sum_{i=1}^{N} I(U_i = K) + a_K \end{aligned}$$

$$(\pi_{kj\eta' c}, \pi_{kj\eta' n}|\mathbf{C}, b_c, b_n) \sim Dirichlet(s_c, s_n)$$
$$\begin{aligned} s_c &= \sum_{i=1}^{N} I(U_i = k, C_{i,j-1} = \eta', C_{ij} = c) + b_c \\ s_n &= \sum_{i=1}^{N} I(U_i = k, C_{i,j-1} = \eta', C_{ij} = n) + b_n \end{aligned}$$

$$\begin{aligned} &P(U_i = k|\mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\alpha}, p_1, \cdots, p_k) \\ &\propto \; p_k \times \left[\prod_{\eta} \omega_{k\eta}(\mathbf{Q}_i)^{I(U_i=k, C_{i1}=\eta)}\right] \left[\prod_{j=2}^{5} \prod_{\eta'} \prod_{\eta} \pi_{kj\eta'\eta}^{I(U_i=k, C_{i,j-1}=\eta', C_{ij}=\eta)}\right] \end{aligned}$$

$$\begin{aligned} &P(C_{ij} = c|Y_{ij}, Z_i, D_{ij}, U_i, \boldsymbol{\lambda}, \mathbf{A}_i, \boldsymbol{\gamma}, \mathbf{W}_i, \boldsymbol{\varphi}_i, \mathbf{Q}_i, \boldsymbol{\alpha}, \sigma^2) \\ &= \begin{cases} \frac{\pi^{**}_{ijc} \times \phi \; \frac{Y_{ij} - (\lambda_{jc0} + \mathbf{A}_i^T \boldsymbol{\gamma} + \mathbf{W}_i^T \boldsymbol{\varphi}_i)}{\sigma}}{\sum_{\eta} \; \pi^{**}_{ij\eta} \times \phi \; \frac{Y_{ij} - (\lambda_{j\eta 0} + \mathbf{A}_i^T \boldsymbol{\gamma} + \mathbf{W}_i^T \boldsymbol{\varphi}_i)}{\sigma}} & \text{if } \in Z_i = 0, D_{ij} = 0, U_i = k \\ 0 & \text{if } \in Z_i = 1, D_{ij} = 0, U_i = k \\ 1 & \text{if } \in Z_i = 1, D_{ij} = 1, U_i = k \end{cases} \end{aligned}$$

$$\begin{aligned} &P(C_{ij} = n|Y_{ij}, Z_i, D_{ij}, U_i, \boldsymbol{\lambda}, \mathbf{A}_i, \boldsymbol{\gamma}, \mathbf{W}_i, \boldsymbol{\varphi}_i, \mathbf{Q}_i, \boldsymbol{\alpha}, \sigma^2) \\ &= \begin{cases} \frac{\pi^{**}_{ijn} \times \phi \; \frac{Y_{ij} - (\lambda_{jn0} + \mathbf{A}_i^T \boldsymbol{\gamma} + \mathbf{W}_i^T \boldsymbol{\varphi}_i)}{\sigma}}{\sum_{\eta} \; \pi^{**}_{ij\eta} \times \phi \; \frac{Y_{ij} - (\lambda_{j\eta 0} + \mathbf{A}_i^T \boldsymbol{\gamma} + \mathbf{W}_i^T \boldsymbol{\varphi}_i)}{\sigma}} & \text{if } \in Z_i = 0, D_{ij} = 0, U_i = k \\ 1 & \text{if } \in Z_i = 1, D_{ij} = 0, U_i = k \\ 0 & \text{if } \in Z_i = 1, D_{ij} = 1, U_i = k \end{cases} \end{aligned}$$

where $\pi^{**}_{ij\eta} = \begin{cases} \prod_{\eta'} \left[\omega_{k\eta}(\mathbf{Q}_i)\pi_{k2\eta\eta'}\right]^{I(U_i=k, C_{i1}=\eta, C_{i2}=\eta')} & \text{if } j = 1 \\ \prod_{\eta'} \prod_{\eta''} \left[\pi_{kj\eta'\eta}\pi_{k,j+1,\eta\eta''}\right]^{I(U_i=k, C_{i,j-1}=\eta', C_{ij}=\eta, C_{i,j+1}=\eta'')} & \text{if } 1 < j < 5 \\ \prod_{\eta'} \pi_{k5\eta'\eta}{}^{I(U_i=k, C_{i4}=\eta', C_{i5}=\eta)} & \text{if } j = 5 \end{cases}$

and $\phi(S)$ is the pdf for standard normal distribution evaluated at S

6 References

ANGRIST, J. D., IMBENS, G. W., AND RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444-455.

BRUCE, M. L., TEN HAVE, T. R., REYNOLDS, C. F., KATZ, I. I., SCHULBERG, H. C., MULSANT, B. H., BROWN, G. K., MCAVAY, G. J., PEARSON, J. L., AND ALEXOPOULOS, G. S. (2004). Reducing suicidal ideation and depressive symptoms in depressed older primary care patients. *Journal of the American Medical Association* **291**, 1081-1091.

COLLINS, L. M., AND WUGALTER, S. E. (1992). Latent class model for stage-sequential dynamic latent variables. *Multivariate Behavioral Research* **27**, 131-157.

FRANGAKIS, C. E., BROOKMEYER, R. S., VARADHAN, R., SAFAEIAN, M., VLAHOV, D., AND STRATHDEE, S. A. (2004). Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program. *Journal of the American Statistical Association* **99**, 239-249.

FRANGAKIS, C. E., AND RUBIN, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86**, 365-379.

FRANGAKIS, C. E., AND RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21-29.

GARRETT, E. S., AND ZEGER, S. L. (2000). Latent class model diagnosis. *Biometrics* **56**, 1055-1067.

GELFAND, A. E., AND SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. (2004). *Bayesian Data Analysis* (2nd ed). New York: Chapman and Hall.

GEMAN, S., AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

GOODMAN, L. A. (1974). Explanatory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215-231.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

HUMPHREYS, K. AND JANSON, H. (2000). Latent transition analysis with covariates, nonresponse, summary statistics and diagnostics: modelling children's drawing development. *Multivariate Behavioral Research* **35**, 89-118.

IMBENS, G. W., AND RUBIN, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. em The Annals of Statistics **25**, 305-327.

LITTLE, R. J., AND RUBIN, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health* **21**, 121-145.

MIGLIORETTI, D. L. (2003). Latent transition regression for mixed outcomes. *Biometrics* **59**, 710-720.

NUIJTEN, M. J., HARDENS, M., SOUETRE, E. (1995). A Markov process analysis comparing the cost effectiveness of maintenance therapy with citalopram versus standard therapy in major depression. *Pharmacoeconomics* **8**, 159-168.

PENG, Y., LITTLE, R. J. A., AND RAGHUNATHAN, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics* **60**, 598-607.

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

REBOUSSIN, B. A., MILLER, M. E., LOHMAN, K. K., AND TEN HAVE, T. R. (2002). Latent class models for longitudinal studies of the elderly with data missing at random. *Applied Statistics* **51**, Part 1, 69-90.

REBOUSSIN, B. A., REBOUSSIN, D. M., LIANG, K. Y., AND ANTHONY, J. C. (1998). Latent transition modeling of progression of health-risk behavior. *Multivariate Behavioral Research* **33**, 457-478.

REVICKI, D. A., BROWN, R. E., PALMER, W., BAKISH, D., ROSSER, W. W., ANTON, S. F., AND FEENY, D. (1994). Modeling the cost-effectiveness of antidepressant treatment in primary-care. *Pharmacoeconomics* **8**, 524-540.

RUBIN, D. B. (1978). Bayesian inference for causal effects. *The Annals of Statistics* **6**, 34-58.

RUBIN, D. B. (1986). Statistics and causal inference. Comment: which ifs have causal answers. *Journal of the American Statistical Association* **81**, 961-962.

TEN HAVE, T. R., ELLIOTT, M. R., JOFFE, M., ZANUTTO, E., AND DATTO, C. (2004). Causal models for randomized physician encouragement trials in treating primary care depression. *Journal of the American Statistical Association* **99**, 16-25.

VELICER, W. F., MARTIN, R. A., AND COLLINS, L. M. (1996). Latent transition analysis for longitudinal data. *Addiction* **91**, S197-S209.

YAU, L. H. Y., AND LITTLE, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* **96**, 1232-1244.

List of Tables

Table 1: Posterior Means and 95% Credible Intervals (in parentheses) for the Time- and Compliance Superclass-Varying Compliance Probabilities Assuming the Average Baseline HAMD of 18.1 and Superclass Probabilities Under the CI Model.

Time	Low Compliers	Decreasing Compliers	High Compliers
4-months	0.43(0.33,0.53)	0.99(0.96,1.00)	1.00(0.98,1.00)
8-months	0.01(0.00,0.07)	0.99(0.94,1.00)	1.00(0.99,1.00)
12-months	0.01(0.00,0.04)	0.51(0.36,0.66)	1.00(0.98,1.00)
18-months	0.06(0.02,0.12)	0.11(0.00,0.28)	0.99(0.98,1.00)
24-months	0.04(0.01,0.09)	0.01(0.00,0.07)	0.83(0.77,0.90)
$P(U_i)$	0.28(0.23,0.33)	0.16(0.12,0.22)	0.56(0.50,0.62)

Table 2: Posterior Means and 95% Credible Intervals (in parentheses) for the ITT Contrasts of the Outcome Within Compliance Superclasses Under the CI Model.

Time	Low Compliers	Decreasing Compliers	High Compliers
4-months	-7.54(-10.05,-2.00)	-1.35(-3.23,0.10)	-1.32(-3.20, 0.09)
8-months	-3.39(- 7.24, 0.81)	-0.93(-2.78,0.83)	-0.92(-2.78, 0.86)
12-months	0.84(- 2.21, 3.95)	-0.61(-2.11,1.05)	-2.03(-3.86,-0.14)
18-months	1.44(- 1.40, 4.07)	1.28(-1.35,3.85)	-1.34(-3.33, 0.64)
24-months	0.04(- 2.58, 2.69)	0.10(-2.61,2.85)	-1.50(-3.72, 0.63)

Table 3: Posterior Means and 95% Credible Intervals (in parentheses) for the Time- and Compliance Superclass-Varying Compliance Probabilities Assuming the Average Baseline HAMD of 18.1 and Superclass Probabilities Under the MCC Model.

Time	Increasing Noncompliers	Erratic Compliers	High Compliers
4-months	0.66(0.53,0.80)	0.38(0.00,1.00)	0.99(0.88,1.00)
8-months	0.38(0.20,0.56)	0.83(0.07,1.00)	0.98(0.86,1.00)
12-months	0.19(0.00,0.40)	0.32(0.00,1.00)	0.99(0.86,1.00)
18-months	0.10(0.02,0.31)	0.93(0.12,1.00)	0.96(0.76,1.00)
24-months	0.02(0.00,0.07)	0.66(0.00,1.00)	0.88(0.65,1.00)
$P(U_i)$	0.42(0.25,0.56)	0.04(0.00.0.15)	0.54(0.42,0.72)

Table 4: Posterior Means and 95% Credible Intervals (in parentheses) of the Transitional Probabilities Under the MCC model.

Superclass	j	$P(C_{i,j} = c \mid C_{i,j-1} = c, U_i)$	$P(C_{i,j} = c \mid C_{i,j-1} = n, U_i)$
Increasing	2	0.57(0.34,0.77)	0.01(0.00,0.06)
Noncomplier	3	0.45(0.00,0.77)	0.01(0.00,0.03)
	4	0.27(0.00,1.00)	0.06(0.02,0.12)
	5	0.10(0.00,0.51)	0.02(0.00,0.05)
Erratic	2	0.67(0.00,1.00)	0.56(0.00,1.00)
Complier	3	0.31(0.00,1.00)	0.48(0.00,1.00)
	4	0.64(0.00,1.00)	0.78(0.00,1.00)
	5	0.68(0.00,1.00)	0.54(0.00,1.00)
High	2	1.00(0.99,1.00)	0.15(0.00,1.00)
Complier	3	1.00(1.00,1.00)	0.44(0.00,1.00)
	4	0.97(0.84,1.00)	0.54(0.00,1.00)
	5	0.91(0.76,1.00)	0.46(0.00,1.00)

Table 5: Posterior Means and 95% Credible Intervals (in parentheses) for the ITT Contrasts of the Outcome Within Compliance Superclasses Under the MCC model.

Time	Increasing Noncompliers	Erratic Compliers	High Compliers
4-months	-5.19(-7.33,-3.04)	-8.32(-15.33,-0.76)	-1.46(-3.05,-0.04)
8-months	-2.70(-5.21,-0.34)	-1.39(- 4.71, 0.58)	-0.89(-2.57, 0.77)
12-months	0.52(-1.92, 3.13)	-0.01(- 3.41, 3.75)	-2.10(-3.81,-0.37)
18-months	1.55(-1.05, 4.23)	-1.28(- 3.29, 1.48)	-1.38(-3.23, 0.50)
24-months	0.48(-2.12, 2.95)	-1.31(- 4.57, 2.35)	-2.02(-4.53, 0.11)