University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2006 *Paper* 10

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

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

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

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

Abstract

This article discusses a nested latent class model for analyzing longitudinal randomized trials when subjects do not always adhere to the treatment to which they are randomized. In the "Prevention of Suicide in Primary Care Elderly: Collaborative Trial" (PROSPECT) study, subjects were randomized to either the control treatment, where they received standard care, or to the intervention, where they received standard care in addition to meeting with depression health specialists. The health specialists educate patients, their families, and physicians about depression and monitor their treatment. Those randomized to the control treatment have no access to the health specialists; however, those randomized to the intervention could choose not to meet with the health specialists, hence, receiving only the standard care. Subjects participated in the study for two years where depression severity and adherence to meeting with health specialists were measured at each follow-up. The outcome of interest is the effect of meeting with the health specialists on depression severity. Traditional "intention-to-treat" and "as-treated" analyses may produce biased causal effect estimates in the presence of subject noncompliance. Utilizing a nested latent class model that uses subject-specific and time-invariant "superclasses" allows us to summarize longitudinal trends of compliance patterns, and estimate the effect of the intervention using "intent-to-treat" contrasts within principal strata corresponding to longitudinal compliance behavior patterns. Analyses show that subjects with more severe depression are more likely to adhere to treatment randomization, and those that are compliant and meet with health specialists benefit from the meetings and show improvement in depression. Simulation results show that our estimation procedure produces reasonable parameter estimates under correct model assumptions.

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

Julia Y. Lin
Thomas R. Ten Have
Michael R. Elliott *

6/21/06

*Julia Y. Lin is a doctoral candidate (E-mail: jlin@cceb.med.upenn.edu), and Thomas R. Ten Have is Professor (E-mail: ttenhave@cceb.med.upenn.edu), Department of Biostatistics and Epidemiology, University of Pennsylvania, School of Medicine, Philadelphia, PA 19104. Michael R. Elliott is Assistant Professor (E-mail: mrelliot@umich.edu), Department of Biostatistics, and Assistant Research Scientist, Institute for Social Research, University of Michigan, Ann Arbor, MI 48109. Julia Y. Lin was supported by grant T32MH065218-1A1, Thomas R. Ten Have was supported by grant R01-MH-61892-01A2, and Michael R. Elliott was supported by grant P30-MH066270. The authors thank the referees for helpful comments.

Abstract

This article discusses a nested latent class model for analyzing longitudinal randomized trials when subjects do not always adhere to the treatment to which they are randomized. In the "Prevention of Suicide in Primary Care Elderly: Collaborative Trial" (PROSPECT) study, subjects were randomized to either the control treatment, where they received standard care, or to the intervention, where they received standard care in addition to meeting with depression health specialists. The health specialists educate patients, their families, and physicians about depression and monitor their treatment. Those randomized to the control treatment have no access to the health specialists; however, those randomized to the intervention could choose not to meet with the health specialists, hence, receiving only the standard care. Subjects participated in the study for two years where depression severity and adherence to meeting with health specialists were measured at each follow-up. The outcome of interest is the effect of meeting with the health specialists on depression severity. Traditional "intention-to-treat" and "as-treated" analyses may produce biased causal effect estimates in the presence of subject noncompliance. Utilizing a nested latent class model that uses subject-specific and time-invariant "superclasses" allows us to summarize longitudinal trends of compliance patterns, and estimate the effect of the intervention using "intent-to-treat" contrasts within principal strata corresponding to longitudinal compliance behavior patterns. Analyses show that subjects with more severe depression are more likely to adhere to treatment randomization, and those that are compliant and meet with health specialists benefit from the meetings and show improvement in depression. Simulation results show that our estimation procedure produces reasonable parameter estimates under correct model assumptions.

KEY WORDS: Longitudinal compliance class model; Principal stratification; Latent class model; Potential outcomes; Randomized trial; Geriatric depression.

1 INTRODUCTION

We consider longitudinal compliance class models that use hierarchical latent class structures which characterize subject compliance behavioral patterns over time. More specifically, we consider studies where subjects are randomized at baseline and the randomization status remains the same over time, though subject adherence to their randomized treatment may vary over time. We illustrate the model with analyses of the "Prevention of Suicide in Primary Care Elderly: Collaborative Trial" (PROSPECT) study (Bruce et al. 2004).

The PROSPECT study was a randomized intervention trial targeted at elderly 64 years of age or older with clinical depression in primary care clinics. There were two treatment groups:

control treatment, where patients received standard care; and intervention, where patients met with health specialists in addition to receiving standard care. The health specialists educated patients, families, and physicians about depression and monitored patient adherence to treatment. Primary care clinics were randomized to the treatment groups rather than randomization of individual patients for practicality and potential contamination between patients in the same clinic. Those randomized to the control treatment did not have access to the health specialists. Patients were randomized to a treatment group only once at baseline, and the randomization status remained constant throughout the study. Patients were followed up at 4, 8, 12, 18, and 24 months, and at each follow-up their adherence to the treatment randomization and outcomes were measured. Treatment adherence was measured by whether patients met with the health specialists at least once since the last follow-up if randomized to the intervention group. The outcome of the study is the severity of patients' depression measured by the Hamilton Rating Scale for Depression (HAMD). The HAMD is a 17-item scale with scores ranging from 0-54. Higher scores indicate more severe depression. In the PROSPECT study, the baseline HAMD scores ranged from 2-41 with a mean of 18.1 and a standard deviation of 6.0. Patient suicidal ideation was measured at baseline and 26% of our study sample had suicidal ideation. We are interested in the effect of the intervention on depression controlling for treatment non-adherence.

Estimating the causal effects of experimental manipulations becomes complicated in the presence of subject noncompliance (Little and Rubin 2000). Traditionally, the intention-to-treat (ITT) and as-treated (AT) analyses are performed but these two methods do not always allow us to obtain unbiased estimates of the causal effects of the treatment on outcomes in the presence of treatment noncompliance. The relationship between experimental manipulation and the outcome may be confounded by treatment non-adherence. Hence, it may be important to account for compliance behavior in the presence of subject noncompliance when estimating treatment effects. Compliers and noncompliers may be inherently different, which could affect how they respond to treatment randomization, as well as the actual treatment itself (Heitjan 1999; Frangakis and Rubin 1999; Mealli, Imbens, Ferro, and Biggeri 2004).

Baker and Lindeman (1994), Angrist, Imbens, and Rubin (1996), and Imbens and Rubin (1997) suggested using latent compliance classes to characterize compliance behavior, within which the effect of treatments are estimated to account for compliance behavior. In a cross-sectional two-arm randomized experimental study, where treatment-received is binary (experimental or control treatment), and subjects have access to the experimental or control treatment regardless of treatment randomization, there are four possible compliance classes: compliers, those that would adhere to the treatment to which they are randomized; always-takers, those that would always seek the experimental treatment regardless of the treatment to which they are randomized; never-takers, those that would always opt for the control treatment regardless of the treatment to which they are randomized; and defiers, those that would seek the treatment opposite to the arm to which they are randomized. Estimating the effect of the experimental treatment while controlling for compliance behavior can be accomplished by estimating the ITT effects within compliance classes via comparing the potential outcomes under each randomization arm stratified on compliance classes. The Imbens and Rubin (1997) approach allows us to examine the effect of the experimental treatment in particular population of interest such as the compliers (i.e. complier average causal effect; CACE), for which the ITT contrast is a valid estimate of the direct effect of treatment on the outcome under certain assumptions (Angrist et al. 1996; Little and Rubin 2000).

Compliance classes are latent variables related to randomization status and actual treatment received. They fall under the principal stratification framework, and can be thought of as "compliance principal strata," such as in Imbens and Rubin (1997), and Frangakis and Rubin (1999, 2002), where principal strata were defined by compliance classes. Knowing the randomization status and the actual treatment received does not completely determine the compliance class. For example, an individual randomized to the control treatment and receives the control treatment could be a complier or a never-taker but not an always-taker or a defier. The probabilities of subject compliance class membership can be estimated in a Bayesian framework (Imbens and Rubin 1997; Frangakis and Rubin 1999; Hirano, Imbens, Rubin, and Zhou 2000), or in a frequentist framework (Little and Yau 1998; Jo and Muthén 2001). Peng, Little, and Raghunathan (2004) compared Bayesian methods to frequentist

methods and concluded that both methods yield similar results.

Not all two-arm cross-sectional studies provide four possible compliance classes. For example, in the PROSPECT study where subjects randomized to the control treatment have no access to the experimental treatment, there are only two possible compliance classes: compliers and never-takers. Since no subjects randomized to the control treatment could be observed to receive the experimental treatment, there are no always-takers and defiers. In this setting, compliance classes are partially latent. Subjects assigned to the treatment arm are either observed to be compliers if they receive the experimental treatment, or never-takers if they do not receive the experimental treatment. However, these compliance classes are not observed for those randomized to the control arm.

We extend the Imbens and Rubin (1997) method to longitudinal settings where treatment is randomized once initially but treatment is applied multiple times throughout the study, and outcomes are repeatedly measured over time. Subject compliance behavior is allowed to vary over time (i.e. treatment received could be different at each follow-up period). We model the relationship between latent, or partially observed, longitudinal compliance patterns and longitudinal outcomes.

There have been other extensions of the Imbens and Rubin's (1997) method to longitudinal settings. Yau and Little (2001) extended Imbens and Rubin's (1997) method to a study where treatment was randomized initially and treatment was applied only once, and outcomes were measured repeatedly over time. Treatment compliance was measured by initial treatment randomization and actual treatment received, and did not vary over time. Frangakis et al. (2004) extended the model to a study where randomization and treatment were administered at multiple times throughout the study, and longitudinal outcomes were measured. Treatment compliance was measured repeatedly over time, and was allowed to vary over time.

The proposed method differs from the method in Yau and Little (2001) in that we allow time-varying treatment compliance. It also differs from the method in Frangakis et al. (2004)

in two ways: 1) we assume that randomization is constant across time; 2) we incorporate a hierarchical latent class structure that consists of time-varying compliance nested in classes of longitudinal compliance trends that are time-invariant. Our method provides summary measures of longitudinal compliance behaviors, and estimates ITT effects within overall compliance behavioral trends.

In longitudinal studies where randomization status remains constant but treatment compliance can vary over time, ITT contrasts must be made with respect to the longitudinal compliance class patterns, which can be thought of as pre-existing, though unobserved, subject-level characteristics at the time of randomization. Thus one might consider the set of potential outcomes for a given compliance pattern, for example, any subject who would have been a complier at every time point. However, it may be useful to characterize longitudinal compliance patterns and estimate the causal effect of treatment conditional on general longitudinal compliance profiles, which we denote as "superclasses". We propose to use latent class models to create subject-specific and time-invariant "superclasses" to describe longitudinal trends of compliance patterns. This allows us to obtain a baseline, time-invariant compliance class structure based on time-varying compliance information that we can utilize in a potential outcomes framework. These "superclasses" become the principal strata (Frangakis and Rubin 2002; Frangakis et al. 2004) within which we estimate the effect of randomization over time. Thus, if the data suggests there are, e.g. "highly compliant," "highly noncompliant," and "early compliant" superclasses, we could consider the effect of randomization among the "highly compliant" as a measure of clinical interest. The time-varying compliance classes in the PROSPECT study are functions of subjects' initial treatment randomization and their actual treatments received at each follow-up, and the superclasses are functions of subjects' compliance classes over time.

We will define notation, discuss assumptions, principal effect of interest, parametric models, parameter estimation, and assessment of model fit in Section 2, and discuss analysis and simulation results in Section 3, then proceed to discussion and future work in Section 4.

2 LONGITUDINAL PRINCIPAL STRATIFICATION MODEL

2.1 Notation

Let Z_i denote randomization for subject i where $i = (1, ..., N)$ and $Z_i \in (0, 1)$ for standard care and the intervention, respectively. Similarly, let D_{ij} denote 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 standard care and the intervention, respectively. Note that Z_i does not have index j because we are restricting to designs in which randomization remains constant over time. Let Y_{ij} denote the observed HAMD score for subject i at time j. We use $\mathbf{Z}$, $\mathbf{D}$, and $\mathbf{Y}$ to denote the vector of Z_i, D_{ij}, and Y_{ij}. Not all subjects have data for all 5 follow-ups, so let J_i denote the last follow-up visit of subject i where $J_i \in (1, 2, 3, 4, 5)$.

Extending the notation of Angrist et al. (1996), we identify two types of potential outcomes: $D_{ij}(Z)$ and $Y_{ij}(Z, D(Z))$. Let $D_{ij}(Z)$ denote the potential treatment received at time j if subject i is randomized to treatment Z; and let $Y_{ij}(Z, D(Z))$ denote the potential outcome for subject i at time j if randomized to group Z and received treatment D(Z). In this paper, we use upper case letters to denote random variables or indices for potential outcomes (e.g. $D_{ij}(Z)$), and lower case letters for realized or observed values of these random variables or potential indices (e.g. $Z_i = z$).

In PROSPECT, patients who are randomized to the control treatment do not have access to the health specialists; therefore, there are only two possible compliance classes: compliers and never-takers. Let C_{ij} denote the latent compliance class for subject i at time j where $C_{ij} \in (c, n)$ for compliers and never-takers, respectively. We use $\mathbf{C}$ to denote the vector of C_{ij}. The potential treatment received, $D_{ij}(0)$ and $D_{ij}(1)$, define compliance class: $C_{ij} = c$ if $D_{ij}(0) = 0$ and $D_{ij}(1) = 1$, and $C_{ij} = n$ if $D_{ij}(0) = 0$ and $D_{ij}(1) = 0$. In studies where those randomized to the control treatment have no access to the intervention, the compliance classes are observed among those randomized to the intervention group: $C_{ij} = c$ if $Z_i = 1$ and $D_{ij} = 1$ and $C_{ij} = n$ if $Z_i = 1$ and $D_{ij} = 0$.

Note that $D_{ij}(Z)$ is a deterministic function of C_{ij}. If we know the compliance class of an individual, then we know the potential treatment received of that individual. Therefore, the notation $(Y_{ij}(Z)|C_{ij})$ is equivalent to $Y_{ij}(Z, D(Z))$ where $Y_{ij}(Z)$ denotes the potential outcome for subject i at time j if randomized to group Z.

In studies such as the PROSPECT where subjects are followed longitudinally and subject compliance behavior could vary over time, the number of possible compliance patterns increases exponentially with time. With two possible compliance classes (i.e. compliers and never-takers) and five time points, we have $2^5 = 32$ possible compliance patterns. We propose to use compliance "superclasses" to summarize compliance patterns present in the data, in which we can stratify on to compare potential outcomes. Let U_i denote the compliance superclass for subject i where $U_i \in (1, ..., K)$, and 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 the probability of compliance class membership, respectively. We use $\mathbf{A}$ and $\mathbf{Q}$ to denote vectors of $\mathbf{A}_i$ and $\mathbf{Q}_i$. These models are discussed in more detail in Section 2.4.

2.2 Assumptions

The randomization assumption and the stable unit-treatment value assumption (SUTVA) are made to identify the potential outcomes and the effect of the treatment using observed outcomes.

The randomization assumption implies that conditional on observed baseline covariates (e.g. baseline HAMD score and suicide ideation), treatment randomization is independent of all baseline variables or pre-randomization variables (e.g. latent baseline characteristics) and potential outcomes (Rubin 1978).

We implicitly make the stable unit-treatment value assumption (SUTVA) of no interference between units when we write $D_{ij}(Z)$ and $Y_{ij}(Z, D(Z))$ instead of $D_{ij}(\mathbf{Z})$ and $Y_{ij}(\mathbf{Z}, \mathbf{D}(\mathbf{Z}))$, where $\mathbf{D}(\mathbf{Z})$ is a vector of potential treatment received for all participants. The SUTVA no

interference assumption says that the compliance behavior of one individual is not affected by randomization status of other individuals, and the potential outcome of one individual is not affected by the randomization and compliance behaviors of other individuals. Note that we are only assuming independence between subjects and do not assume independence within subjects.

Identification of potential outcomes also require SUTVA's consistency assumption (Rubin 1986). It assumes that the potential outcome of a certain treatment will be the same regardless of the treatment assignment mechanism. This assumption may be violated when there are version of a treatment not reflected in the treatment indicator variable. The consistency assumption implies that the potential outcome $Y_{ij}(Z = z, D(Z = z))$ for observed z is the observed outcome conditional on $Z_i = z$, and the potential treatment received $D_{ij}(Z = z)$ for observed z is the observed treatment received conditional on $Z_i = z$. Under this assumption, the observed outcome is a function of the potential outcomes and the observed randomization status, and the observed treatment received is a function of the potential treatment received and observed randomization status: $Y_{ij} = Z_i * Y_{ij}(1, D(1)) + (1 - Z_i) * Y_{ij}(0, D(0))$ and $D_{ij} = Z_i * D_{ij}(1) + (1 - Z_i) * D_{ij}(0)$.

2.3 Principal Effects

If we attempt to estimate ITT effect within each of the 32 possible combinations of the time-varying compliance classes in the PROSPECT study, we may encounter the problem of having little or no subjects in some of the patterns, and inference based on such estimates would be highly sensitive to modelling assumptions. We would also have 32 ITT effects for the different compliance patterns, which could make interpretation of the results difficult, and may be of limited use clinically.

An alternative to stratifying on the time-varying compliance classes is to stratify on "compliance superclasses" that characterize longitudinal compliance patterns. The superclasses serve as summary measures of compliance patterns within which we estimate ITT effects of

the experimental treatment. A compliance superclass is a pre-randomization variable, which allows us to model potential outcomes, just as we do with compliance classes. It is a latent variable at the subject level; unlike the time-varying compliance classes as we defined here, which is at both subject and time level. The effect of interest would be the principal effect of treatment assignment on the outcome at time j within a compliance superclass:

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

This principal effect is a standard ITT contrast, although stratified on the latent compliance superclass, which allows us to consider the effect of randomization on the outcome controlling for subject longitudinal compliance behavior. Since the compliance superclasses do not represent exact longitudinal compliance class patterns, this principal effect does not translate to the CACE in Angrist et al. (1996). We discuss details of the superclass model in Section 2.4.

The principal effect of interest can be identified through the observed data by making the randomization and the SUTVA consistency assumptions:

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

The first equality follows from the randomization assumption which assumes independence between baseline characteristics, such as potential outcomes, and the randomization status conditional on other baseline factors, such as U_i. The second equality follow from SUTVA's consistency assumption, which assumes that the observed outcomes are functions of the potential outcomes and the observed randomization status. Assuming the randomization and the SUTVA consistency assumptions, we can utilize observed outcomes to model potential outcomes.

2.4 Parametric Models

Consider the complete data that consist of randomization status, potential outcomes, and potential treatment received. Let $f(.)$ denote the distribution function. The complete data distribution is given by:

$$\begin{aligned} & f\left(Z_1, \cdots, Z_N, Y_{11}(Z=0), \cdots, Y_{NJ_N}(Z=0), Y_{11}(Z=1), \cdots, Y_{NJ_N}(Z=1),\right. \\ & \quad \left. D_{i1}(Z=0), \cdots, D_{NJ_N}(Z=0), D_{i1}(Z=1), \cdots, D_{NJ_N}(Z=1)\right) \qquad (1) \\ = & \prod_{i=1}^{N} f(Z_i=z) f\left(Y_{i1}(Z=0), \cdots, Y_{iJ_i}(Z=1), D_{i1}(Z=0), \cdots, D_{iJ_i}(Z=1)\right) \end{aligned}$$

The equality follows from the randomization and the SUTVA no interference assumptions. The distribution of the observed data $(Y_{i1}, \cdots, Y_{iJ_i}, D_{i1}, \cdots, D_{iJ_i})$ for subject i is given by:

$$\begin{aligned} & f(Y_{i1}, \cdots, Y_{iJ_i}, D_{i1}, \cdots, D_{iJ_i}|\boldsymbol{\theta}) \\ = & \int\int f(Y_{i1}(Z=0), \cdots, Y_{iJ_i}(Z=1), D_{i1}(Z=0), \cdots, D_{iJ_i}(Z=1)|\boldsymbol{\theta})\, dY_i^{mis}\, dD_i^{mis} \\ = & \int\int f(Y_{i1}(Z=0), \cdots, Y_{iJ_i}(Z=1)|D_{i1}(Z=0), \cdots, D_{iJ_i}(Z=1), \boldsymbol{\theta}) \\ & \quad f(D_{i1}(Z=0), \cdots, D_{iJ_i}(Z=1)|\boldsymbol{\theta})\, dY_i^{mis}\, dD_i^{mis} \\ = & \int\int f(Y_{i1}(Z=0), \cdots, Y_{iJ_i}(Z=1)|C_{i1}, \cdots, C_{iJ_i}, U_i, \theta^Y) \qquad (2) \\ & \quad f(C_{i1}, \cdots, C_{iJ_i}|U_i, \theta^C) f(U_i|\theta^U)\, dY_i^{mis}\, dD_i^{mis} \\ = & \int\int f(Y_{i1}(Z=0), \cdots, Y_{iJ_i}(Z=1)|C_{i1}, \cdots, C_{iJ_i}, \theta^Y) \\ & \quad f(C_{i1}, \cdots, C_{iJ_i}|U_i, \theta^C) f(U_i|\theta^U)\, dY_i^{mis}\, dD_i^{mis} \end{aligned}$$

where Y_i^{mis} and D_i^{mis} contains the $Y_{ij}(Z)$ and $D_{ij}(Z)$ for which $Z \neq Z_i$, and $\boldsymbol{\theta} = (\theta^Y, \theta^C, \theta^U)$ where θ^Y parameterizes the distribution of the potential outcomes $Y_{ij}(Z)$, and θ^C and θ^U parameterize the latent compliance class and superclass model for C_{ij} and U_i.

Following Hirano et al. (2000) and Imbens and Rubin (1997), we model the outcome conditional on partially latent compliance class, randomization, and covariates. We consider a hierarchical model structure, similar to Elliott, Gallo, Ten Have, Bogner, and Katz (2005)

and Hogan and Daniels (2002), to account for subject-level random effect. Outcomes are assumed to be conditionally independent given compliance classes $C_{i1}, \cdots, C_{ij}$, treatment assignment Z_i, subject-level baseline covariates $\mathbf{A}_i$, and subject-level random effect $\boldsymbol{\varphi}_i$:

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

Let $\boldsymbol{\lambda}$ denotes the vector of $\lambda_{t\eta' z}$. The within-compliance class effect of randomization on the outcome is represented by $\sum_{t=1}^{j}\left[\sum_{\eta'} I(C_{it} = \eta', Z_i = z)\lambda_{t\eta' z}\zeta(t,j)\right]$, where $I(H)$ is an indicator function that equals to 1 if H is true, and 0 otherwise. We use $\zeta(t,j)$ to describe the relationship between compliance class membership at time t on the outcome at time j, where $t \leq j$. If we assume a transient relationship, where outcome is assumed to depend only on the compliance behavior in the current time period, then $\zeta(t,j) = I(t = j)$; if we assume a non-transient relationship, where outcome is assumed to dependent on compliance behavior in the current and all prior time periods, then $\zeta(t,j) = I(t \leq j)$; if we assume a decaying relationship, where outcome is assumed to be dependent on compliance behavior in the current and all prior time periods, but the effect diminishes as time lag increases, then we may assume $\zeta(t,j) = e^{-\tau(j-t)}$ where $\tau > 0$.

The fixed effect of subject-level baseline covariates $\mathbf{A}_i$ on the outcome are described by the column vector $\boldsymbol{\gamma}$. The random effect $\boldsymbol{\varphi}_i$ is used to account for the within-subject correlation in the outcomes where $\mathbf{W}_i$ denotes the random effect design matrix of subject i. A preliminary analysis indicated that the within-clinic correlation was small (0.075); hence we ignore it for this analysis. Subject-level random effects are assumed to be independent of random errors of the outcome, and of the latent classes C_{ij} and U_i for all i and j.

We assume that the compliance superclass is an underlying factor that drives subject compliance over time; therefore, we model the time-varying compliance class conditional on compliance superclass. Following Hirano et al. (2000) and Ten Have, Elliott, Joffe, Zanutto, and Datto (2004), we model the compliance class probabilities conditional on baseline covariates

$\mathbf{Q}_i$. Let $P(C_{ij} = \eta | U_i = k, \mathbf{Q}_i) = \pi_{kj\eta}(\mathbf{Q}_i)$, and $\pi_{kj\eta}(\mathbf{Q}_i) = exp(\alpha_{0kj\eta} + \boldsymbol{\alpha}_{1\eta}\mathbf{Q}_i)/[\sum_{\eta'} exp(\alpha_{0kj\eta'} + \boldsymbol{\alpha}_{1\eta'}\mathbf{Q}_i)]$,where $\sum_{\eta'} \pi_{kj\eta'}(\mathbf{Q}_i) = 1 \ \forall k, j$ (Clogg 1995), and we constrain $\alpha_{0kj\eta^*}$ and $\boldsymbol{\alpha}_{1\eta^*}$ for one of the compliance class η^* to be 0 for identifiability. This multinomial logit model assumes a constant effect of the baseline covariates $\mathbf{Q}_i$ on the compliance class probabilities that does not vary by time nor compliance superclass. Compliance classes between subjects and compliance classes within a subject at each time point conditional on compliance superclass and baseline covariates are assume to be independent:

$$P(\mathbf{C}|\mathbf{U}, \mathbf{Q}) = \prod_{i=1}^{N} P(C_{i1}, ..., C_{iJ_i}|U_i, \mathbf{Q}_i) = \prod_{i=1}^{N} P(C_{i1}|U_i, \mathbf{Q}_i)...P(C_{iJ_i}|U_i, \mathbf{Q}_i).$$

Assuming conditional independence between compliance classes given superclass, estimates of the $\boldsymbol{\lambda}$ parameters do not change with different $\zeta(t, j)$. We prove this in the Appendix. This allows us to use the transient conditional mean outcome model in analysis, where $\zeta(t, j) = I(t = j)$, regardless of the true relationship between the compliance classes and the outcomes. Thus, equation (3) becomes $\mu_{ijz} = I(C_{ij} = \eta, Z_i = z)\lambda_{j\eta z} + \mathbf{A}_i^T\boldsymbol{\gamma} + \mathbf{W}_i^T\boldsymbol{\varphi}_i$.

We assume that subject compliance superclass $(U_i = k) \sim Multinomial(1, p_k)$, where $\sum_{k=1}^{K} p_k = 1$. Compliance superclasses between subjects are assumed to be independent: $f(\mathbf{U}) = \prod_{i=1}^{N} f(U_i = k)$ for $k = 1, ..., K$. We assume that there are no covariates associated with compliance besides the covariates $\mathbf{Q_i}$ used in modelling time-varying compliance, and therefore, do not include any covariates in modelling the superclasses. Covariates can be incorporated using multinomial logit models.

Given the model specifications, Equation (2) becomes

$$
\begin{aligned}
&\int\int f(Y_{i1}(Z=0),\cdots,Y_{iJ_i}(Z=1)|C_{i1},\cdots,C_{iJ_i},\theta^Y) \\
&\qquad f(C_{i1},\cdots,C_{iJ_i}|U_i,\theta^C)f(U_i|\theta^U)\,dY_i^{mis}\,dD_i^{mis} \\
&= \int\int f(Y_{i1}(Z=0),Y_{i1}(Z=1)|C_{i1},\theta^Y)\cdots f(Y_{iJ_i}(Z=0),Y_{iJ_i}(Z=1)|C_{iJ_i},\theta^Y) \\
&\qquad f(C_{i1}|U_i,\theta^C)\cdots f(C_{iJ_i}|U_i,\theta^C)f(U_i|\theta^U)\,dY_i^{mis}\,dD_i^{mis} \\
&= \left[\prod_{j=1}^{J_i}\sum_{k'=1}^{K} p_{k'}\left(\begin{array}{l}\pi_{k'jc}(\mathbf{Q}_i)\times\phi\left(\frac{Y_{ij}-(\lambda_{jc0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)+\\ \pi_{k'jn}(\mathbf{Q}_i)\times\phi\left(\frac{Y_{ij}-(\lambda_{jn0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)\end{array}\right)\right]^{I(Z_i=0)} \times \\
&\qquad \left[\prod_{j\in D_{ij}=1}\sum_{k'=1}^{K} p_{k'}\ \phi\left(\frac{Y_{ij}-(\lambda_{jc1}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)\right]^{I(Z_i=1)} \times \\
&\qquad \left[\prod_{j\in D_{ij}=0}\sum_{k'=1}^{K} p_{k'}\ \phi\left(\frac{Y_{ij}-(\lambda_{jn1}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)\right]^{I(Z_i=1)}
\end{aligned}
$$

where $\phi(s)$ is the pdf for standard normal distribution evaluated at s.

2.5 Estimation

We use the Bayesian Markov Chain Monte Carlo (MCMC) method to estimate model parameters. The model set up is similar to the methods used in Hirano et al. (2000) and Ten Have et al. (2004). Bayesian estimation of latent compliance classes proposed by Imbens and Rubin (1997) requires specifying prior distribution. We assume $\boldsymbol{\alpha} \sim MVN\,(\mathbf{0}, \Sigma_\alpha)$ (Garrett and Zeger 2000; Ten Have et al. 2004), and the prior $(p_1, ..., p_K) \sim Dirichlet(a_1, ..., a_K)$. 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 $\mathbf{X}_i$ denotes the design matrix of the fixed effect for subject i with J_i rows. Let $\boldsymbol{\beta} = [\lambda_{1c0}, \cdots, \lambda_{5n1}, \boldsymbol{\gamma}]$ denote the vector of coefficients corresponding to the fixed effect. We put a priors on the vector of fixed effect parameters $\boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}_\beta, \Sigma_\beta)$, the random effect parameters, $\boldsymbol{\varphi}_i \sim MVN\,(\mathbf{0}, \Sigma_\varphi)$, and conjugate hyperprior, $\Sigma_\varphi \sim Inv - Wishart\,(df = \nu_\varphi, \Gamma)$. We also assume a conjugate prior for the variance of the random errors, $\sigma^2 \sim Inv - \chi^2\,(df = \nu_\sigma, \psi)$. Gibbs sampling (Geman and

Geman 1984; Gelfand and Smith 1990; Imbens and Rubin 1997; Gelman, Carlin, Stern, and Rubin 2004) is used to obtain draws from the posterior distributions of the parameters. The detailed description of the posterior distributions of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\varphi}_i$, Σ_φ, C_{ij}, p_k, and U_i from which the parameters are drawn are given in the Appendix.

Since the posterior distributions of $(\boldsymbol{\alpha}|\mathbf{C}, \mathbf{U}, \mathbf{Q})$ is not of a known parametric form, we propose to use the Metropolis-Hasting algorithm (Hastings 1970; Gelman et al. 2004) technique to draw the $\boldsymbol{\alpha}$ parameters. We use the multivariate t-distribution with 3 degrees of freedom as the proposal distribution, using estimates of $\boldsymbol{\alpha}$ from the previous iteration and the inverse of the observed information as the mean and variance, respectively.

2.6 Model Fit Assessment

To assess the fit of the model to the data we examine the posterior predictive distributions (PPD; Gelman et al. 2004). Since the compliance classes in those randomized to the intervention are observed, we compare the PPD of the time-varying compliance classes in those randomized to the intervention and assess their fit to the observed compliance. Let G_m denote the number of individuals in 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:

$$S^{obs} = \sum_m \frac{(G_m^{obs} - N\kappa_m)^2}{N\kappa_m(1-\kappa_m)} \text{ and } S^{rep} = \sum_m \frac{(G_m^{rep} - N\kappa_m)^2}{N\kappa_m(1-\kappa_m)} \tag{4}$$

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

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

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

3 RESULTS

In the PROSPECT study, 320 subjects were in clinics that were randomized to intervention, and 278 were in clinics that were randomized to the standard care. All subjects at each time point that have not dropped out of the study and have HAMD measures are considered for the analysis. At the 4-month follow-up, 234 remained of those randomized to standard care and had HAMD measures, and 253 remained of those randomized to the intervention and had HAMD measures. Of those 253 who were randomized to the intervention, 30 received the standard care instead, and no one randomized to the standard care was observed to have received the intervention as the intervention would not have been accessible to them. At the 24-month follow-up, 151 were left of those randomized to the standard care and had HAMD measures, and 170 remained of those randomized to the intervention and had HAMD measures. Of those 170, 48 received the standard care instead. We assume that dropouts and missing outcomes are occurring at random (MAR; Little and Rubin 2002). We include the baseline HAMD score and baseline suicidal ideation in modelling the outcome because we are interested in the change in HAMD from baseline in each subject and because randomization failed to balance proportions of subjects with suicidal ideation between the control and the intervention groups at baseline (Bruce et al. 2004). ITT analysis and AT analysis will be performed in addition to analysis based on the superclass model for comparison.

3.1 ITT analysis

In an ITT analysis, we ignore the treatment that subjects actually received, and look at the effect of randomization on the outcome. In this analysis, we fit a linear mixed effects model with random subject-level intercepts, and we control for baseline HAMD score and baseline suicidal ideation. Table 1 shows that the effect of being randomized to meet with health specialists is strongest during the first four months of the study with an average of a 3-point reduction in the HAMD score. The effect diminishes until the 24-month follow-up where we see an increase in reduction of HAMD score in those that were randomized to meet the

health specialists relative to those that were randomized to the usual care.

An ITT effect can provide biased estimates of the direct effect of the intervention when subject noncompliance is present and not accounted for in the analysis. It may not reflects the effect of the intervention because the subjects' randomization status is not necessarily the same as the actual treatment received.

3.2 AT analysis

In an AT analysis, we ignore the subject randomization status, and look at the effect of the actual intervention received on the outcome. In this AT analysis, we also fit a linear mixed effects model controlling for baseline HAMD score and baseline suicidal ideation, with subject-level random intercepts. The AT analysis result is shown in Table 1. The effect of actually meeting the health specialists is the strongest during the first four months of the study. The effect diminishes until the 24-month follow-up where we see an increase in reduction of HAMD scores in those who met with the health specialists relative to those who did not. In contrast to the ITT effect, the AT effect is significant over the entire observation period.

AT estimates of the effect of the intervention may also be biased in the presence of noncompliance or confounders related to the treatments and the outcome (Heitjan 1999). When we ignore the randomization, the groups may no longer be comparable.

3.3 Superclass Model

In this section we consider the application of our extension of the Imbens and Rubin (1997) method to a longitudinal case where randomization stays constant over time but the treatment received could vary. We consider a transient model, where outcomes are only dependent on current compliance behavior and not prior compliance behavior. Under the conditional independence assumption for the compliance classes given compliance superclass, estimates

obtained under the transient model are equal to those obtained under a non-transient model as shown in the Appendix. The effect of interest is the principal effect of treatment assignment on the outcome within a compliance superclass controlling for baseline HAMD score and baseline suicidal ideation:

$$E[Y_{ij}(Z_i = 1)|U_i = k] - E[Y_{ij}(Z_i = 0)|U_i = k] = \sum_{\eta'} \pi_{kj\eta'}(\mathbf{Q}_i) \times (\lambda_{j\eta'1} - \lambda_{j\eta'0}) \qquad (5)$$

In PROSPECT, those randomized to the usual care have no access to the health specialists; hence, there are only two possible compliance classes: complier and never-taker. We include baseline HAMD score in modelling the compliance probabilities.

In our superclass model we use priors that are not informative for the Bayesian MCMC algorithm because we do not have strong a priori inclinations. We assume
$\boldsymbol{\alpha} \sim MVN\left(\mathbf{0}, \Sigma_\alpha = diag(9/4, 0.04)\right)$ for intercepts and coefficient for the effect of baseline HAMD score on compliance class probabilities following Garrett and Zeger (2000) and Ten Have et al. (2004). The difference in the variance component in the prior reflects the different scaling of the covariates. A larger variance is used for binary covariates (intercept) and a smaller variance is used for continuous covariates (baseline HAMD score). We assume a noninformative prior $(p_1, ..., p_K) \sim Dirichlet(1, ..., 1)$ for the compliance superclass probabilities. We assume a diffuse prior on the coefficients for the within-compliance class effect of randomization, baseline HAMD score, and baseline suicidal ideation on the outcomes, $\boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}_\beta = \mathbf{0}, \Sigma_\beta = 1000 \times \mathbf{I})$. In this analysis, we assume a model with subject-level intercepts and assume the conjugate prior $\Sigma_\varphi \sim Inv{-}\chi^2\left(\nu_\varphi = 1, \Gamma = 1/10\right)$. We also assume a conjugate prior for the variance of the random errors, $\sigma^2 \sim Inv - \chi^2\left(\nu_\sigma = 1, \psi = 1/10\right)$. Identifiability of the $\boldsymbol{\alpha}$ parameters is checked by comparing the prior and posterior distributions (Garrett and Zeger 2000).

We assume there are three compliance superclasses. Goodman (1974) suggests that we can only identify at most 3 latent compliance superclasses given we have dichotomous compliance classes for 5 time points.

We run 3 MCMC chains of 5,000 iterations after discarding 5,000 iterations of burn-in. To deal with possible label switching of the latent classes from one iteration to another, we use the algorithm proposed by Stephens (2000) to ensure that the labels of latent classes in one iteration is equivalent to the labels in other iterations. To assess the convergence of the chains, we calculate the $\hat{R}$ statistic (Gelman et al. 2004, pp.296-297) which is the square root of the ratio between the weighted average of the within- and between-chain variance and the within-chain variance. $\hat{R} = 1$ is considered convergence, and $\hat{R} < 1.1$ is acceptable; our maximum value of $\hat{R}$ was 1.04.

Table 2 shows the compliance probabilities at each follow-up within each compliance superclasses given an average HAMD of 18.1. We see that the probability of membership in the complier class is low in the first superclass, and it decrease over the first year then stabilizes in the second year. We will call this superclass "low compliers". In the second superclass, the probability of compliance starts high then rapidly decreases over time. We will call this superclass "decreasing compliers." In the third superclass, the probability of compliance is near 1, and remains high for the first 18 months, then drops off slightly in the last 6 months. We will call this last superclass "high compliers." The posterior probabilities and their associated 95% credible interval of membership in the low compliers, decreasing compliers, and high compliers superclasses are 0.19(0.12,0.26), 0.07(0.00,0.14), and 0.74(0.67,0.80), respectively.

The posterior mean and 95% credible interval for the log odds of complier class membership with an one-point increase in baseline HAMD score is 0.04(-0.004,0.08). It suggests that patients with more severe depression at baseline, indicated by higher baseline HAMD scores, are more likely to be compliers than never-takers compared to those with less severe depression at baseline.

Table 3 presents the posterior means and credible intervals of (5), the ITT effect of randomization on the outcome within each compliance superclass controlling for baseline HAMD and baseline suicidal ideation. The ITT contrast within the low compliers shows a large decrease in the HAMD score at 4 months, suggesting that those in clinics randomized to the intervention show greater improvement in depression than those in clinics randomized to

the usual care. Never-takers receive the usual care regardless of their randomization status; therefore, the ITT effect at 4 months in the low complier superclass is likely to be largely contributed by the direct effect of randomization. We also see a reduction in HAMD scores in the ITT contrasts within decreasing compliers and high compliers, who are highly compliant at 4 month, suggesting that meeting the health specialists when randomized to the intervention may be effective in improving the status of depression. At 24 months, we see the largest reduction in HAMD in the ITT contrast of the high complier superclass. These findings suggest that although being in clinics randomized to the intervention, or having access to the health specialists, may help improve depression initially, subjects need to actually meet with the health specialists to benefit from the intervention longitudinally.

We compare the fit of the posterior predictive distribution (PPD) of the longitudinal compliance class patterns in those randomized to the intervention under the two-superclass model to that under the three-superclass model. The PPD p-value for the two-superclass model is 0.003, and is 0.046 for the three-superclass model. This gives evidence that the three-superclass model has a better fit to the data than the two-superclass model.

3.4 Simulation Study

We simulated 100 data sets of 400 subjects each with complete 5 follow-up visits assuming there are two superclasses: increasing-noncompliers, whose compliance probabilities are 0.50, 0.40, 0.30, 0.20, and 0.10 from the first visit to the fifth visit; steady-compliers, whose compliance probabilities are 0.90 throughout all the visits. We assume equal probabilities of being increasing-noncompliers and steady-compliers. We also assume the variances $\sigma^2 = 22$ and $\Sigma_\varphi = 25$. We simulated the data under $\boldsymbol{\beta} = (5.84, 2.79, 5.23, 2.38, 3.65, 2.12, 3.51,$
$1.65, 2.46, 1.49, 3.56, 2.21, 1.96, 1.16, 2.73, 3.51, 1.70, -0.04, 2.38, 2.43, 0.44)$ which corresponds to ITT effects in the increasing-noncompliers as -2.950, -1.728, -1.236, 0.464, and -0.129 from the fist visit to the last visit, and ITT effects in the steady-compliers as -3.030, -1.563, -1.008, -0.642, and -1.561. We analyze the simulated data assuming the correct model structure.

The means of the estimated compliance probabilities, the compliance superclass probabilities, and the ITT effect within each superclass of the simulated data sets are presented in Table 4. Our estimation procedure seems to perform well in estimating the compliance probabilities and the compliance superclass probabilities. The estimated values are very close to the true values from which the data is simulated. The estimated ITT effects within superclasses also seem reasonable relative to the true values. Table 5 presents the coverage of the 95% credible intervals of the estimates. It shows good coverage for the compliance probability and ITT effect estimates. We see an over-coverage of the compliance superclass probability estimates. A binary variable is expected to have the largest variance when the probability of "success" is at 0.5. Therefore, we expect the variance of the compliance superclass probabilities to be the largest when the true probability of being in one of two classes is 0.5.

4 DISCUSSION

The proposed model extends the methods in Imbens and Rubin (1997) and Yau and Little (2001) to longitudinal data where randomization is constant through time but the compliance behavior may vary over time. Similar to Frangakis et al. (2004), the proposed model accommodates time-varying latent classes. However, unlike Frangakis et al. (2004), we do not allow randomization to vary over time. Another difference between the proposed model and the model in Frangakis et al. (2004) is that we utilize a nested latent class structure to summarize longitudinal time-varying compliance behaviors and estimate principal effects within the time-invariant subject-specific compliance superclasses. The proposed method allows us to make ITT contrasts controlling for longitudinal compliance behaviors by stratifying on the compliance superclasses. If we were to compare outcomes stratified on specific patterns of compliance over time, it may result in too many comparisons to be of any practical use, or we may run into problems with small sample sizes in some of the patterns. Furthermore, the Imbens and Rubin (1997) framework allows us to assess the direct effect of randomization on the outcome. Although utilizing the superclasses provide convenient summaries of the longitudinal compliance patterns, we lose the ability to have clear causal

interpretations of the stratified ITT contrasts.

In the PROSPECT study, we found a relationship between baseline depression severity and compliance behaviors. Those with more severe depression at baseline, indicated by higher HAMD score, are more likely to be compliers than never-takers than those with less severe depression. It is sensible that those with more severe depression would be more cooperative in working with their physicians to improve their condition, and be more willing to adhere to the treatment to which they are randomized. Those with less severe depression may not feel the need for treatment; hence, reject intervention when it is available to them.

We identified three types of longitudinal compliance profiles in the analysis: 1) low compliance; 2) decreasing compliance and 3) high compliance. The ITT contrasts at each time point within superclasses suggest that there may be a direct effect of randomization initially, but those who are more compliant and met with health specialists when randomized to the intervention showed greater improvement in depression than those who are less compliant at the end of the study. This suggests that meeting with health specialists is effective in improving depression. Although those who did not comply when randomized to the intervention still showed reduction in the HAMD score early in the study, those who complied were able to benefit from meeting with the health specialists and showed greater reduction in the HAMD score longitudinally. An advantage of our approach over a cross-sectional ITT analysis, and over even a cross-sectional principal stratum analysis, is that we were able to identify a latent class of highly compliant individuals whose ITT effects at later time periods were greater than those estimated through conventional approaches, consistent with a hypothesis of treatment benefit.

The proposed nested latent class model also allows us to examine the direct effect of randomization. In groups with large proportion of never-takers, such as in our low complier superclass, the ITT effects are likely due to the direct effect of randomization rather than the effect of the actual intervention. In the PROSPECT study, we found evidence of direct effect of randomization early on but the effect diminished over time. In large clinical randomization studies, clustered randomization are often performed due to concerns of contamination

between subjects in the same cluster randomized to different treatment groups. Our result suggests that with adequate run-in time for the intervention, the effect of randomization becomes minimal. It may be acceptable to randomize individual subjects within clusters given sufficient follow-up time.

The simulation results show that our estimation method produces reasonable estimates with generally good coverage when the underlying assumptions are correct. The over-coverage of the compliance superclass probability estimates seems to indicate that our estimates may be too conservative.

In our analysis we assume three compliance superclasses, the maximum number of latent classes that can be identified when the dichotomous "outcomes" (in this case the compliance class membership) are available (Goodman 1974). However, our model fit analysis using PPD-probabilities showed that while model fit for the 3-superclass model (p=0.046) was better than for the 2-superclass model (p=0.003), it was still less than ideal. Hence we are currently exploring Markov Chain models for the compliance classes as a richer model than the "conditional independence" model considered here.

Beyond relaxing the conditional independence assumption of compliance classes across time given superclass and covariates, we can also consider incorporating other covariates in modelling the compliance classes and superclasses to identify better predictors of compliance. If clinicians can identify patients who are likely to comply to treatment over time and those less likely to comply, then they may be able to target patients with particular attributes or tailor treatment for different patients to optimize treatment adherence and treatment outcomes.

The reviewers raised an important issue of the effect of prior outcomes on compliance behaviors. In this analysis we model the joint distribution of the compliance classes and the outcomes as a product of the joint distribution of compliance classes and joint distribution of the outcomes conditional on the joint distribution of the compliance classes:

$$f(Y_{i1}, \cdots, Y_{iJ_i}, C_{i1}, \cdots, C_{iJ_i}) = f(Y_{i1}, \cdots, Y_{iJ_i} | C_{i1}, \cdots, C_{iJ_i}) f(C_{i1}, \cdots, C_{iJ_i})$$

This allows us to incorporate the relationship between compliance classes and prior outcomes,

though this relationship is not modelled explicitly. Alternatively we can model the joint distributions as a time series:

$$
\begin{aligned}
& f(Y_{i1}, \cdots, Y_{iJ_i}, C_{i1}, \cdots, C_{J_i}) \\
&= P(C_{i1})P(Y_{i1}|C_{i1})P(C_{i2}|Y_{i1}, C_{i1})P(Y_{i2}|C_{i1}, C_{i2}, Y_{i1}) \cdots \\
&\quad P(C_{iJ_i}|Y_{i1}, \cdots, Y_{i,J_i-1}, C_{i1}, \cdots, C_{i,J_i-1})P(Y_{iJ_i}|C_{i1}, \cdots, C_{iJ_i}, Y_{i1}, \cdots, Y_{i,J_i-1})
\end{aligned}
$$

However, examining this model raised the issue of identifiability in a potential outcomes framework, and it is not clear how one would construct meaningful ITT contrasts in this setting. Nonetheless, it would be valuable to explore the relationship between prior outcomes and compliance behaviors in future work.

5 Appendix

5.1 Proof within-superclass ITT effects are the same under the transient, non-transient, and decay models assuming compliance classes at each time points are conditionally independent from each other given compliance superclass.

We consider transient, non-transient, and decay models. We are interested in the ITT contrast within superclass.

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

Let $\lambda^*_{j\eta z}$ be the compliance class-specific effect of randomization on the outcome for compliance class η at time j when randomized to treatment group z under the transient model, and $\lambda_{j\eta z}$ be the compliance class-specific effect of randomization under the non-transient model.

ITT contrast at time j under transient model:

$$P(C_{ij}=c|U_i=k)(\lambda^*_{jc1}-\lambda^*_{jc0}) + P(C_{ij}=n|U_i=k)(\lambda^*_{jn1}-\lambda^*_{jn0})$$

ITT contrast at time j under non-transient model:

$$\begin{aligned}
&P(C_{i1}=c|U_i=k)(\lambda_{1c1}-\lambda_{1c0}) + P(C_{i1}=n|U_i=k)(\lambda_{1n1}-\lambda_{1n0}) + ... \\
&+P(C_{ij}=c|U_i=k)(\lambda_{jc1}-\lambda_{jc0}) + P(C_{ij}=n|U_i=k)(\lambda_{jn1}-\lambda_{jn0}) \\
&= [P(C_{i1}=c|U_i=k)\lambda_{1c1} + P(C_{i1}=n|U_i=k)\lambda_{1n1} + ... + \\
&\quad P(C_{ij}=c|U_i=k)\lambda_{jc1} + P(C_{ij}=n|U_i=k)\lambda_{jn1}] \qquad (6) \\
&- [P(C_{i1}=c|U_i=k)\lambda_{1c0} + P(C_{i1}=n|U_i=k)\lambda_{1n0} + ... + \\
&\quad P(C_{ij}=c|U_i=k)\lambda_{jc0} + P(C_{ij}=n|U_i=k)\lambda_{jn0}]
\end{aligned}$$

$$
\begin{aligned}
&P(C_{i,j-1}=c|U_i=k)\lambda_{j-1,cz}+P(C_{i,j-1}=n|U_i=k)\lambda_{j-1,nz}\\
&= [P(C_{i,j-1}=c,C_{ij}=c|U_i=k)+P(C_{i,j-1}=c,C_{ij}=n|U_i=k)]\,\lambda_{j-1,cz}\\
&+ [P(C_{i,j-1}=n,C_{ij}=c|U_i=k)+P(C_{i,j-1}=n,C_{ij}=n|U_i=k)]\,\lambda_{j-1,nz}\\
&= [P(C_{i,j-1}=c|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i,j-1}=c|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{j-1,cz}\\
&+ [P(C_{i,j-1}=n|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i,j-1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{j-1,nz}
\end{aligned}
$$

assuming conditional independence of compliance probabilities given superclass.

$$
\begin{aligned}
&P(C_{i1}=c|U_i=k)\lambda_{1c1}+P(C_{i1}=n|U_i=k)\lambda_{1n1}\\
&= [P(C_{i1}=c,C_{i2}=c,...,C_{ij}=c|U_i=k)+...+P(C_{i1}=c,C_{i2}=n,...,C_{ij}=n|U_i=k)]\,\lambda_{1c1}\\
&+ [P(C_{i1}=n,C_{i2}=c,...,C_{ij}=c|U_i=k)+...+P(C_{i1}=n,C_{i2}=n,...,C_{ij}=n|U_i=k)]\,\lambda_{1n1}
\end{aligned}
$$

summing over all 2^{j-1} compliance patterns for C_{i2} to C_{ij}.

$$
\begin{aligned}
&= [P(C_{i1}=c,C_{i2}=c,...,C_{i,j-1}=c|U_i=k)P(C_{ij}=c|U_i=k)+...+\\
&\quad P(C_{i1}=c,C_{i2}=n,...,C_{i,j-1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1c1}\\
&+ [P(C_{i1}=n,C_{i2}=c,...,C_{i,j-1}=c|U_i=k)P(C_{ij}=c|U_i=k)+...+\\
&\quad P(C_{i1}=n,C_{i2}=n,...,C_{i,j-1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1n1}\\
&= [P(C_{i1}=c|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i1}=c|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1c1}\\
&+ [P(C_{i1}=n|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1n1}
\end{aligned}
$$

$$
\begin{aligned}
&P(C_{i1}=c|U_i=k)\lambda_{1c0}+P(C_{i1}=n|U_i=k)\lambda_{1n0}\\
&= [P(C_{i1}=c,C_{i2}=c,...,C_{ij}=c|U_i=k)+...+P(C_{i1}=c,C_{i2}=n,...,C_{ij}=n|U_i=k)]\,\lambda_{1c0}\\
&+ [P(C_{i1}=n,C_{i2}=c,...,C_{ij}=c|U_i=k)+...+P(C_{i1}=n,C_{i2}=n,...,C_{ij}=n|U_i=k)]\,\lambda_{1n0}\\
&= [P(C_{i1}=c,C_{i2}=c,...,C_{i,j-1}=c|U_i=k)P(C_{ij}=c|U_i=k)+...+\\
&\quad P(C_{i1}=c,C_{i2}=n,...,C_{i,j-1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1c0}\\
&+ [P(C_{i1}=n,C_{i2}=c,...,C_{i,j-1}=c|U_i=k)P(C_{ij}=c|U_i=k)+...+\\
&\quad P(C_{i1}=n,C_{i2}=n,...,C_{i,j-1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1n0}\\
&= [P(C_{i1}=c|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i1}=c|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1c0}\\
&+ [P(C_{i1}=n|U_i=k)P(C_{ij}=c|U_i=k)+P(C_{i1}=n|U_i=k)P(C_{ij}=n|U_i=k)]\,\lambda_{1n0}
\end{aligned}
$$

Then equation (6) becomes:

$$
\begin{aligned}
&P(C_{ij}=c|U_i=k)\,[P(C_{i1}=c|U_i=k)\lambda_{1c1}+P(C_{i1}=n|U_i=k)\lambda_{1n1}+\\
&\qquad P(C_{i2}=c|U_i=k)\lambda_{2c1}+P(C_{i2}=n|U_i=k)\lambda_{2n1}+...+\\
&\qquad P(C_{i,j-1}=c|U_i=k)\lambda_{j-1,c,1}+P(C_{i,j-1}=n|U_i=k)\lambda_{j-1,n,1}+\lambda_{jc1}]\\
&+P(C_{ij}=n|U_i=k)\,[P(C_{i1}=c|U_i=k)\lambda_{1c1}+P(C_{i1}=n|U_i=k)\lambda_{1n1}+\\
&\qquad P(C_{i2}=c|U_i=k)\lambda_{2c1}+P(C_{i2}=n|U_i=k)\lambda_{2n1}+...+\\
&\qquad P(C_{i,j-1}=c|U_i=k)\lambda_{j-1,c,1}+P(C_{i,j-1}=n|U_i=k)\lambda_{j-1,n,1}+\lambda_{jn1}]\\
&-P(C_{ij}=c|U_i=k)\,[P(C_{i1}=c|U_i=k)\lambda_{1c0}+P(C_{i1}=n|U_i=k)\lambda_{1n0}+\\
&\qquad P(C_{i2}=c|U_i=k)\lambda_{2c0}+P(C_{i2}=n|U_i=k)\lambda_{2n0}+...+\\
&\qquad P(C_{i,j-1}=c|U_i=k)\lambda_{j-1,c,0}+P(C_{i,j-1}=n|U_i=k)\lambda_{j-1,n,0}+\lambda_{jc0}]\\
&-P(C_{ij}=n|U_i=k)\,[P(C_{i1}=c|U_i=k)\lambda_{1c0}+P(C_{i1}=n|U_i=k)\lambda_{1n0}+\\
&\qquad P(C_{i2}=c|U_i=k)\lambda_{2c0}+P(C_{i2}=n|U_i=k)\lambda_{2n0}+...+\\
&\qquad P(C_{i,j-1}=c|U_i=k)\lambda_{j-1,c,0}+P(C_{i,j-1}=n|U_i=k)\lambda_{j-1,n,0}+\lambda_{jn0}]\\
&=P(C_{ij}=c|U_i=k)\lambda^*_{jc1}+P(C_{ij}=n|U_i=k)\lambda^*_{jn1}\\
&-P(C_{ij}=c|U_i=k)\lambda^*_{jc0}-P(C_{ij}=n|U_i=k)\lambda^*_{jn0}\\
&=P(C_{ij}=c|U_i=k)(\lambda^*_{jc1}-\lambda^*_{jc0})+P(C_{ij}=n|U_i=k)(\lambda^*_{jn1}-\lambda^*_{jn0})
\end{aligned}
$$

Therefore, when fitting transient model to non-transient data, we still get unbiased ITT effect within superclass. A similar proof shows that fitting transient model to data from decay model would still give us unbiased ITT effect within superlcass.

5.2 Conditional draws of the Gibbs sampler

Let $\mathbf{Y}_i$, denote the column vector of Y_{ij} for subject i. For 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 J_i rows. Let $\boldsymbol{\beta} = [\lambda_{1c0}, \cdots, \lambda_{5n1}, \boldsymbol{\gamma}]$ denote the vector of coefficients corresponding to fixed effect.

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

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

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

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

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

In studies where the control group have no access to the intervention, such as the PROSPECT study, the only possible compliance classes are compliers and never-takers. Therefore $\pi_{kjc}(\mathbf{Q}_i) = exp(\alpha_{0kjc} + \boldsymbol{\alpha}_{1c}\mathbf{Q}_i)/[1 + exp(\alpha_{0kjc} + \boldsymbol{\alpha}_{1c}\mathbf{Q}_i)]$ and $\pi_{kjn}(\mathbf{Q}_i) = 1/[1 + exp(\alpha_{0kjc} + \boldsymbol{\alpha}_{1c}\mathbf{Q}_i)]$. The

posterior probabilities of compliance class membership is:

$$P(C_{ij}=c|Y_{ij},Z_i,D_{ij},U_i,\boldsymbol{\lambda},\mathbf{A}_i,\boldsymbol{\gamma},\mathbf{W}_i,\boldsymbol{\varphi}_i,\mathbf{Q}_i,\boldsymbol{\alpha},\sigma^2)$$
$$=\begin{cases}\dfrac{\pi_{kjc}(\mathbf{Q}_i)\times\phi\ \ \frac{Y_{ij}-(\lambda_{jc0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}}{\sum_{\eta'}\left[\pi_{kj\eta'}(\mathbf{Q}_i)\times\phi\ \ \frac{Y_{ij}-(\lambda_{j\eta'0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)\right]} & \text{if } \in Z_i=0,D_{ij}=0,U_i=k\\ 0 & \text{if } \in Z_i=1,D_{ij}=0,U_i=k\\ 1 & \text{if } \in Z_i=1,D_{ij}=1,U_i=k\end{cases}$$

$$P(C_{ij}=n|Y_{ij},Z_i,D_{ij},U_i,\boldsymbol{\lambda},\mathbf{A}_i,\boldsymbol{\gamma},\mathbf{W}_i,\boldsymbol{\varphi}_i,\mathbf{Q}_i,\boldsymbol{\alpha},\sigma^2)$$
$$=\begin{cases}\dfrac{\pi_{kjn}(\mathbf{Q}_i)\times\phi\ \ \frac{Y_{ij}-(\lambda_{jn0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}}{\sum_{\eta'}\left[\pi_{kj\eta'}(\mathbf{Q}_i)\times\phi\ \ \frac{Y_{ij}-(\lambda_{j\eta'0}+\mathbf{A}_i^T\boldsymbol{\gamma}+\mathbf{W}_i^T\boldsymbol{\varphi}_i)}{\sigma}\right)\right]} & \text{if } \in Z_i=0,D_{ij}=0,U_i=k\\ 1 & \text{if } \in Z_i=1,D_{ij}=0,U_i=k\\ 0 & \text{if } \in Z_i=1,D_{ij}=1,U_i=k\end{cases}$$

The posterior probabilities of compliance superclass is:

$$(p_1,...,p_K|\mathbf{U},a_1,...,a_K)\sim Dirichlet(r_1,...,r_K)$$
$$r_1=\textstyle\sum_1^N I(U_i=1)+a_1$$
$$r_K=\textstyle\sum_1^N I(U_i=K)+a_K$$

Let $\mathbf{C}_i$ denote the vector of C_{ij} for subject i, and $\boldsymbol{\alpha}$ denote the vector of $\alpha_{0kj\eta}$ and $\boldsymbol{\alpha}_{1\eta}$. The posterior probability of compliance superclass membership is:

$$P(U_i=k|\mathbf{C}_i,\mathbf{Q}_i,\boldsymbol{\alpha},p_1,...,p_K)\propto p_k\times\prod_{j=1}^{J_i}\prod_{\eta'}\pi_{kj\eta'}(\mathbf{Q}_i)^{I(C_{ij}=\eta',U_i=k)}$$

Estimation of the parameters is a MCMC process. Note that $\lambda_{j\eta z}$ is a function of the compliance class, C_{ij}; and $\alpha_{0kj\eta}$ is a function of the compliance superclass, U_i. The parameters $\lambda_{j\eta z}$, $\alpha_{0kj\eta}$, and $\boldsymbol{\alpha}_{1\eta}$ are updated at each iteration with the latest estimates of C_{ij} and U_i, which are then used to estimate C_{ij} and U_i in the next iteration.

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.

Baker, S. G., and Lindeman, K. S. (1994), "The Paired Availability Design: A Proposal For Evaluating Epidural Analgesia During Labor," *Statistics in Medicine*, 13, 2269-2278.

Bruce, M. L., Ten Have, T. R., Reynolds, C. F., Katz, I. R., Schulberg, H. C., Mulstant, 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.

Clogg, C. C. (1995), "Latent Class Models," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, eds. G. Arminger, C. C. Clogg, and M. E. Sobel, New York: Plenum Press.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., and Katz, I. R. (2005), "Using a Bayesian Latent Growth Curve Model to Identify Trajectories of Positive Affect and Negative Events Following Myocardial Infarction," *Biostatistics*, 6, 119-143.

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.

— (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). "Exploratory 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.

Heitjan, D. F. (1999), "Ignorability and Bias In Clinical Trials," *Statistics in Medicine*, 18, 2421-2434.

Hirano K., Imbens, G. W., Rubin, D. B., and Zhou, X. (2000), "Assessing the Effect of an Influenza Vaccine in an Encouragement Design," *Biostatistics*, 1, 69-88.

Hogan, J. W., and Daniels, M. J. (2002), "A Hierarchical Modelling Approach to Analysing Longitudinal Data With Drop-out and Non-compliance, With Application to an Equivalence Trial in Paediatric Acquired Immune Deficiency Syndrome," *Applied Statistics*, 51, Part 1, 1-21.

Imbens, G. W. and Rubin, D. B. (1997), "Bayesian Inference for Causal Effects in Randomized Experiments With Noncompliance," *The Annals of Statistics*, 25, 305-327.

Jo, B., and Muthén, B. (2001), "Longitudinal Studies With Intervention and Noncompliance: Estimation of Causal Effects in Growth Curve Mixture Modeling," in *Multilevel*

Modeling: Methodological Advances, Issues, and Applications, eds. N. Duan and S. P. Reise, New York: Lawrence Erlbaum Associates, pp. 51-62.

Little, R. J. and Rubin, D. B. (2000), "Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concept and Analytical Approaches," *Annual Reviews of Public Health*, 21, 121-145.

— (2002), *Statistical Analysis With Missing Data* (2nd ed.), New York: John Wiley and Sons Inc.

Little, R. J., and Yau, L. H. Y. (1998), "Statistical Techniques for Analyzing Data From Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model," *Psychological Methods*, 3, 2, 147-159.

Mealli, F., Imbens, G. W., Ferro, S., and Biggeri, A. (2004), "Analyzing A Randomized Trial On Breast Self-Examination With Noncompliance and Missing Outcomes," *Biostatistics*, 5, 207-222.

Peng, Y., Little, R. J. A., and Raghunathan, T. E. (2004), "An Extended General Location Model for Causal Inference from Data Subject to Noncompliance and Missing Values," *Biometrics*, 60, 598-607.

Rubin, D. B. (1978), "Bayesian Inference For Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34-58.

— (1986), "Statistics and Causal Inference. Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, 81, 961-962.

Stephens, M. (2000), "Dealing With Label Switching In Mixture Models," *Journal of the Royal Statistical Society*, Series B, 62, 795-809.

Ten Have, T. R., Elliott, M. R., Joffe, M., Zanutto, E., and Datto, C. (2004), "Causal Models for Randomized Physician Encouragement Trials in Treating in Primary Care Depression," *Journal of the American Statistical Association*, 99, 16-25.

Yau, L. H. Y., 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.

7 TABLES

Table 1. ITT and AT Estimates of Treatment Effect and Corresponding 95% Confidence Intervals (in parentheses).

Time	ITT Effects	AT Effects
4-months	-2.97 (-4.20,-1.74)	-2.54 (-3.67,-1.41)
8-months	-1.58 (-2.84,-0.32)	-1.22 (-2.39,-0.04)
12-months	-1.32 (-2.63,-0.01)	-1.31 (-2.53,-0.09)
18-months	-0.66 (-1.98, 0.66)	-1.31 (-2.55,-0.07)
24-months	-1.42 (-2.83,-0.01)	-2.10 (-3.44,-0.77)

Table 2. Posterior Means and 95% Credible Intervals (in parentheses) for the Time- and Compliance Superclass-Varying Compliance Probabilities Given Average HAMD 18.1 and Compliance Superclass Probabilities.

Time	Low Compliers	Decreasing Compliers	High Compliers
4-months	0.37(0.13,0.57)	0.85(0.26,0.99)	0.99(0.97,1.00)
8-months	0.14(0.02,0.35)	0.80(0.23,0.99)	0.99(0.97,1.00)
12-months	0.08(0.02,0.21)	0.51(0.13,0.96)	0.99(0.96,1.00)
18-months	0.13(0.04,0.28)	0.33(0.03,0.94)	0.98(0084,1.00)
24-months	0.14(0.03,0.34)	0.22(0.02,0.92)	0.89(0.82,0.95)
$P(U_i = k)$	0.19(0.12,0.26)	0.07(0.00,0.14)	0.74(0.67,0.80)

Table 3. Posterior Means and 95% Credible Intervals (in parentheses) for the ITT Effect of Intervention Within Compliance Superclasses.

Time	Low Compliers	Decreasing Compliers	High Compliers
4-months	-9.00(-12.58,-6.14)	-3.19(-10.45,-0.70)	-1.53(-2.87,-0.16)
8-months	-4.47(- 8.66, 0.17)	-1.76(- 2.86, 0.16)	-0.88(-2.41, 0.71)
12-months	1.69(- 4.47, 5.56)	0.05(- 2.89, 3.23)	-1.84(-3.42,-0.11)
18-months	1.71(- 1.75, 5.01)	1.13(- 2.50, 4.69)	-1.23(-2.84, 0.38)
24-months	-0.47(- 3.54, 2.54)	-0.44(- 3.30, 2.89)	-1.62(-3.32,-0.02)

Table 4. Mean Estimated Compliance Probabilities, Compliance Superclass Probabilities, and ITT Effect of Intervention Within Compliance Superclasses (true value in parentheses) From 100 Simulations.

	Increasing-Noncompliers		Steady-Compliers	
Visit	$P(C_{ij} = c)$	Effect	$P(C_{ij} = c)$	Effect
1	0.51(0.5)	-2.99(-2.95)	0.89(0.9)	-2.83(-3.03)
2	0.40(0.4)	-1.60(-1.73)	0.90(0.9)	-1.37(-1.56)
3	0.30(0.3)	-0.80(-1.24)	0.89(0.9)	-1.27(-1.01)
4	0.20(0.2)	0.91(0.46)	0.89(0.9)	-0.89(-0.64)
5	0.11(0.1)	0.21(-0.13)	0.88(0.9)	-1.79(-1.56)
$P(U_i = k)$	0.50(0.50)		0.50(0.50)	

Table 5. Coverage of 95% Credible Intervals for Estimated Compliance Probabilities, Compliance Superclass Probabilities, and ITT Effect of Intervention Within Compliance Superclasses From 100 Simulations.

	Increasing-Noncompliers		Steady-Compliers	
Visit	$P(C_{ij} = c)$	Effect	$P(C_{ij} = c)$	Effect
1	92%	98%	95%	98%
2	97%	99%	96%	99%
3	98%	98%	94%	98%
4	93%	100%	95%	99%
5	97%	99%	92%	99%
$P(U_i = k)$	100%		100%	