Biometrics 000, 000–000 DOI: 000
000 0000

Compliance Mixture Modelling with a Zero Effect Complier Class and Missing Data.

Michael E. Sobel[1,*]

Bengt Muthén[2]

[1] Department of Statistics, Columbia University New York, NY 10027

[2] Professor Emeritus, University of California, Los Angeles

*email: michael@stat.columbia.edu

Summary: Randomized experiments are the gold standard for evaluating proposed treatments. The intent to treat estimand (ITT) measures the effect of treatment assignment, but not the effect of treatment if subjects take treatments to which they are not assigned. The desire to estimate the efficacy of the treatment in this case has been the impetus for a substantial literature on compliance over the last 15 years. In papers dealing with this issue, it is typically assumed there are different types of subjects, for example, those who will follow treatment assignment (compliers), and those who will always take a particular treatment irrespective of treatment assignment. The estimands of primary interest are the complier proportion and the complier average treatment effect (CACE). To estimate CACE, researchers have used various methods, for example, instrumental variables and parametric mixture models, treating compliers as a single class. However, it is often unreasonable to believe all compliers will be affected. This paper therefore treats compliers as a mixture of two types, those belonging to a zero effect class, others to an effect class. Second, in most experiments, some subjects drop out or simply do not report the value of the outcome variable, and the failure to take into account missing data can lead to biased estimates of treatment effects. Recent work on compliance in randomized experiments has addressed this issue by assuming missing data are missing at random or latently ignorable. We extend this work to the case where compliers are a mixture of types and also examine alternative types of non-ignorable missing data assumptions.

Key words: Causal Inference; Compliance; Latent Ignorability; Missing Data; Mixture Distributions

1. Introduction

Randomized experiments are the gold standard for evaluating proposed treatments. Researchers design these treatments to affect the outcome through mediators that lie on one or more pathways to the outcome.

Several questions are generally of interest. First, what is the average effect of offering the treatment on the outcome? This, the average of the heterogeneous unit effects of treatment assignment, is the intent to treat (ITT) estimand. If all subjects report the value of the outcome, this question can be answered by comparing outcomes of subjects in the treatment and control groups. If the ITT is expected to vary over sub-populations of interest, outcomes can be compared within sub-populations.

Second, researchers often want to know the efficacy of the treatment, that is, the actual effect of receiving treatment. This can differ from the ITT if subjects do not take up treatment when assigned to do so or take up treatment when not assigned to do so (Bloom 1984) and/or if treatment assignment affects the outcome "directly". Sorting out the contributions of efficacy and non-compliance to the ITT is important. If the ITT is "small" and the preponderance of subjects comply with their assignments, either the theory behind the intervention is weak and/or the intervention fails to implement the theory well. However, if a treatment has a large effect on the outcome, but subjects assigned to the treatment group do not take the treatment, the estimated ITT may underestimate treatment efficacy. Here, the underlying theory may be sound, but the program delivery may require modification.

Building on Angrist, Imbens and Rubin (1996), a large statistical literature has developed, in which the primary goal is to estimate the so-called "Complier Average Causal Effect" (CACE). The idea is that there are different types of subjects: 1) never takers, who never take the treatment (even when assigned to do so) 2) always takers, who always take the treatment, (even when it is not assigned) and 3) compliers, who will not take the treatment

when they are not assigned to treatment, and who will take the treatment when assigned to do so. By assumption, there are no subjects who will not take the treatment when it is assigned and who will take the treatment when it is not assigned. The ITT is thus a weighted average of the ITT's for the three groups. For type 1) and 2) subjects, the experiment provides no information on the effect of treatment, as these subjects receive the same treatment in both study arms. Still, there may be a "direct" effect of offering the treatment in these groups. However, if this is nil, the ITT is zero for these subjects; the ITT is then the product of the CACE and the (non-zero) proportion of subjects who are compliers. The CACE is then the instrumental variable estimand (Angrist et al. 1996), and in a randomized study it can be estimated by the difference between the treatment and control group means, divided by the difference in the proportion of subjects who take up treatment in the treatment and control arms. If the "direct" effect of offering the treatment is also 0 for the compliers, the CACE is then the average of the heterogeneous unit effects of receiving treatment in this subgroup.

Subsequently, Imbens and Rubin (1997) showed that under the assumptions above the complier marginal distributions under treatment and no treatment are identified, and they showed how to use this information to obtain improved estimates of CACE. More recently, Cheng, Qin and Zhang (2009) used empirical likelihood to propose a semiparametric instrumental variable estimator for the complier marginal distributions.

Little and Yau (1998) incorporated covariates. In their application, using data from the Job Search Intervention Study (JOBS-II, Vinokur, Price and Shul 1995), subjects in the control group did not have access to treatment. In this case, common in clinical trials and prevention studies, we refer to subjects who would take treatment when assigned as compliers. Conditional on covariates, they assumed that outcomes in the treatment and control groups for the different subject types followed normal distributions with common error variance, and under the "exclusion restriction" that the direct effect of treatment on

the outcome is 0 for never takers, the conditional means for the never takers are identical in the treatment and control groups. The probability of being a complier was also allowed to depend on covariates. Maximum likelihood was used to estimate the model parameters.

Subsequent researchers extended the work above in various directions. Hirano, Imbens, Rubin and Zhou (2000) modeled a binary outcome using Bayesian inference without imposing the exclusion restriction among never takers and always takers. Jin and Rubin (2008) addressed the case of partial compliance using principal stratification (Frangakis and Rubin 2002). Small, Ten Have, Joffe and Cheng (2006) modeled repeated binary outcomes with time varying compliance. Walker, White and Babiker (2004) consider survival data.

This paper pursues a different line. For never or always taker subjects, the effect of treatment assignment on receipt of treatment is 0, and if assignment does not directly affect the outcome, the effect of assignment on the outcome is 0. The remaining subjects (assuming there are no defiers) are compliers. In this group, substantive investigators are often interested in knowing the treatment effect and the proportion of subjects who do not respond to treatment. For example, Kowatch et al. (2000) studied the effects of several drugs (lithium, divalproex sodium, carbamezepine) on bipolar disorder in adolescents; the subject's score on the Young Mania Rating Scale was the outcome. They classified compliers in the treatment groups who did not achieve a "sufficient" reduction in the outcome between baseline and final evaluation as non-responders; the remaining compliers were then compared with subjects in the control group to estimate the effect of the drug. They concluded that divalproex sodium had the largest effect; further, while less than 40 percent of subjects treated with lithium (carbamezepine) "responded" to treatment, more than half the subjects treated with divalproex sodium did so. Notwithstanding some problems, the standard analysis above suggests a) there are different classes of compliers, a "zero effect class" in which treatment does not affect the outcome and an effect class, in which the effects may vary across subjects

(as in the case of a drug that affects most subjects positively, a few negatively), and b) the ability to distinguish two classes is often of interest.

In medical and behavioral research, an intervention may be designed to affect a specific mediating variable (variables) hypothesized to in turn affect the outcome. For never and always takers, the intervention does not affect the mediator(s). For compliers, the mechanism(s) targeted by the intervention may be unaffected and/or the mediator(s) does (do) not affect the outcome for some fraction of the subjects. For such subjects, either the effect of the intervention on the mediator(s) is 0 and/or the effect of the mediator(s) on the outcome is 0, and if the direct effect of assignment on the outcome is also 0, the effect of assignment on the outcome is 0. These subjects constitute a "zero effect complier class". The remaining subjects constitute a"complier effect class". Here the effects may vary across subjects, as in the case of a drug that affects most subjects positively, a few negatively.

For example, in the JOBS-II study, the subjects, who were unemployed and looking for work, were assigned to attend/not attend multiple training sessions to learn job search skills. Six months later, depression status was assessed. Almost half the subjects are never takers; for the remaining compliers, it seems reasonable to believe that attending increased search skills, even if only moderately. But while better search skills may have affected depression for some compliers by helping them to become re-employed or at least feel more hopeful about obtaining employment, others may not have been depressed due to their job situation and/or had other issues that prevented them from becoming re-employed or feeling their chances of re-employment were good. For these subjects, enhanced search skills are unlikely to have affected depression.

More generally, distinguishing between subject types is of practical and theoretical interest. If the proportion of always takers is "large", there is little need for the intervention. A "large" never taker proportion could result from offering the intervention to subpopulations

of subjects who do not stand to benefit and/or believe they do not stand to benefit, and/or it could result from poor program delivery. Additional information would be required to distinguish these possibilities.

If the complier proportion is deemed substantial, it will also be of value to distinguish the different complier types. While a large CACE is indicative of a successful intervention, knowing the relative proportions in the zero and effect classes adds useful information. If the zero class is small, one might want to implement the intervention as is, or with minor modifications. If the zero class is relatively large, this suggests improving the delivery to affect the mediators within the zero class and/or adding program components to target other pathways to the outcome; to maket a more focused recommendation, additional information on the mediators and/or prior knowledge would be required. A small CACE is consistent with a small zero effect class and a small effect in the effect class or a relatively large zero class and a somewhat larger effect in the effect class. The first case could be due to weak delivery, but if this possibility can be ruled out, the treatment is not very effective, suggesting consideration of other treatments. Similarly, if the zero class is large and the effect in that class is still "small", that may also suggest consideration of other treatments. If the zero class is large and the effect in the effect class is deemed "large", this suggests improving the delivery and/or adding components to target other pathways, as in the case of a large CACE and a large zero class.

Non-response is an additional concern in randomized intervention studies and the failure to address this issue can lead to biased estimates (Frangakis and Rubin 1999). Yau and Little (2001) extended the analysis in Little and Yau (1998) to a longitudinal setting, allowing missing data on the outcome and compliance status to be missing at random (MAR), given treatment status, treatment assignment and covariates. Frangakis and Rubin (1999) also considered randomized experiments where subjects in the control group cannot receive

treatment. They assumed the missing data are "latently ignorable", meaning the missing data would be MAR if it were possible to observe whether a subject is a never-taker or a complier. They then estimated CACE assuming latent ignorability and the "compound exclusion" restriction that never taker responses and outcomes are unaffected by treatment assignment. Peng, Little and Raghunathan (2004) extended this approach to include covariates. In their study of breast self examination, Mealli, Imbens, Ferro and Biggeri (2004) argued that it is more reasonable to assume that treatment assignment has no effect on non-response for the compliers and they estimated the CACE under this restriction, the exclusion restriction that treatment assignment has no effect on never taker outcomes, and the latent ignorability assumption. Chen, Geng and Zhou (2009) studied the identification and estimation of CACE with a binary outcome and missingness related to the outcome.

This paper also considers randomized experiments in which control group subjects cannot access the treatment. We make several contributions. First, we split the compliers into an effect class and a zero effect class. Using parametric mixture models, we estimate the complier average causal effect in the effect class (hereafter ECACE) and the proportions in the two complier classes and the never taker class. Next, we extend the analysis to the case with missing outcome data. We estimate ECACE and the class proportions under an extension of the latent ignorability assumption (herafter ELI) and either the assumption that the response behavior of zero effect class compliers is the same as that of other compliers or the assumption that the response behavior of zero effect class compliers is the same as that of never takers. In the special case where there is no zero effect complier class, the two class case considered in previous literature is obtained; here, neither of the assumptions just described nor the exclusion restriction that never taker (Frangakis and Rubin 1999) or complier (Mealli et al. 2004)) response probabilities do not depend on treatment assignment is needed.

The paper proceeds as follows. Section two describes the JOBS-II data used to illustrate

our approach and introduces the notation used here. Section three sets out identifying assumptions for the case with no missing data and compares our three class model with the two class model fitted by Little and Yau (1998). Additional assumptions about the response process are introduced in section 4 and used to extend the three class model to the case with missing outcome data. We then examine the performance of the model under alternative missing data assumptions. Section 5 concludes.

2. The JOBS-II Study

2.1 *Data*

The JOBS II study was conducted to ascertain the effect of a job search intervention program on the re-employment and depression of unemployed workers, especially those at high risk for depression. The treatment consisted of 5 half day seminars to teach job search strategies. In addition, treatment group members were given a job search pamphlet. Control group members received only the pamphlet.

Subjects were recruited from offices of the Michigan Employment Security Commission. Persons indifferent to their treatment assignment, out of work for less than 13 weeks, looking for work and not intending to retire within the next two years were eligible to participate. These persons were sent a baseline questionnaire measuring demographic and social psychological variables, including depression level. Respondents to the questionnaire were randomized. Depression levels were reassessed two weeks, six months and two years after the seminars were conducted.

The subject's change in depression six months after intervention (section 3) and the subject's depression six months after intervention (section 4) are the outcome variables. Depression is measured using 11 items from the Hopkins Symptom Checklist (Derogatis, Lipman, Rickles, Uhlemuth, and Covi (1974). The covariates in the analyses include demo-

graphic variables collected at baseline (age, years of education completed, a binary variable for marital status (0 if married, 1 otherwise), income, and a binary variable for race (1 if black, 0 otherwise), a measure of economic hardship and several psychological variables: 1) depression at baseline, 2) subject's motivation to attend, measured using a two item scale, and 3) assertiveness, measured on a four point scale.

Using a risk score described in Price, van Ryn and Vinokur (1992), Little and Yau (1998), Jo (2002) and Yau and Little (2001) split subjects into a group at high risk for depression and a group at low risk, conducting separate analyses in the two groups. The risk score is also used as a covariate in the analyses in section 3.

The analyses herein focus on the high risk group of 715 subjects. To illustrate the utility of introducing a zero effect compliance class, in section three we reanalyze a subset of 502 cases with no missing data used by Little and Yau (1998), comparing our three class model with their two class model. In section four, we additionally allow for missing data on the outcome, analyzing the subset of 657 cases with complete covariate data. Here, as we are not comparing our results to those obtained by previous workers, we use depression at six months as the outcome, including baseline depression as a covariate. Estimation is carried out by maximum-likelihood using an accelerated EM algorithm as discussed in Muthén and Shedden (1999) and Muthén and Asparouhov (2009); all analyses use the Mplus computer program (Muthén and Muthén, 1998-2011).

2.2 *Notation*

For subject $i = 1, ...n$, we observe $Z_i = 0$ if i is assigned to the control group, 1 otherwise, and $\mathbf{X}_i$, a vector of pretreatment covariates. Throughout, we make the standard assumptions that each subject can be assigned to receive treatment or not and the "no interference assumption" that subject's outcomes, response behaviors and treatment take-up do not depend upon the assignments to which other subjects can be allocated (Rubin 1980).

For $z = 0, 1$, let $Y_i(z)$ denote the value of subject i's outcome under treatment assignment z, let $R_i(z) = 1$ if i reports the outcome $Y_i(z)$ under treatment assignment z, 0 otherwise, and let $D_i(z) = 1$ if subject i takes up treatment under assignment z, 0 otherwise. In this study, as subjects in the control group cannot access the treatment, $D_i(0) = 0$ for all i. Subject type is indexed by U, where $D_i(1) = 0$ if and only if subject i is a never-taker ($U_i = 1$), and $U_i = 2$ (effect class compliers) or $U_i = 3$ (zero effect class compliers) implies $D_i(1) = 1$.

The values $Y_i(z), D_i(z), R_i(z)$ cannot be observed for the treatment assignment that i was not assigned. Thus only $Y_i = Y_i(Z_i) = Z_iY_i(1)+(1-Z_i)Y_i(0)$ can be observed; similar remarks apply to R_i and D_i. The observed data are thus $\{(D_i, R_i, Z_i, \mathbf{X}_i, (Y_i : R_i = 1), i = 1, ..., n)\}$; the observations are assumed to be independent and identically distributed, given treatment assignment and the covariates $\mathbf{X}$. Throughout, we assume that assignment is randomized:

A1 (RANDOM ASSIGNMENT)

$$Y(0), Y(1), R(0), R(1), X, U \perp\!\!\!\perp Z, \tag{1}$$

where the symbol $\perp\!\!\!\perp$ denotes statistical independence.

3. Compliance Mixture Modelling with no Missingness

To illustrate the basic approach and compare it with previous work (Little and Yau 1998, Jo 2002), in this section we proceed as if there were no missing data, that is, $R_i = 1$ for all i.

3.1 *Estimands of Interest*

The primary estimands are the densities $f(y(z) \mid u, \mathbf{x}, \theta_{uz}^{UZ})$, depending on parameters θ_{uz}^{UZ}, z = 0,1, u= 1,2,3, and the class proportions $\Pr(U = u \mid \mathbf{X} = \mathbf{x}, \lambda) = \pi_u(\mathbf{x}, \lambda)$, depending on parameters λ. From these, the ECACE $\mathrm{E}(Y(1) - Y(0) \mid U = 2)$, and the class proportions $\pi(u) = \Pr(U = u)$, are obtained.

We use maximum likelihood to model the distribution of Y and D, given $\mathbf{X}$ and Z. For

$d = 0, 1$, $r = 0, 1$, $z = 0, 1$, let S_{drz}^{DRZ} denote the set of observations with D=d, R=r,Z=z. Using assumption A1, the likelihood is

$$L(\xi, \lambda) = \prod_{d,z} \prod_{i \in S_{d1z}^{DRZ}} f(y_i, d_i \mid z_i, \mathbf{x}_i, \theta_{dz}^{DZ}, \lambda) = \prod_{d,z} \prod_{i \in S_{d1z}^{DRZ}} f(y_i(z_i) \mid d_i, z_i, \mathbf{x}_i, \theta_{dz}^{DZ}) f(d_i(z_i) \mid \mathbf{x}_i, \lambda), \quad (2)$$

where $\theta_{dz}^{DZ} = \{\theta_{uz}^{UZ} : (U = u, Z = z) \Rightarrow (D = d, Z = z)\}$, and $\theta = (\theta_{10}'^{UZ}, ..., \theta_{31}'^{UZ})' = K\xi$, where ξ is the vector of distinct elements of θ.

To estimate ξ and λ, additional assumptions are necessary. Assumption A2 is typically made in randomized studies with imperfect compliance:

A2 (EXCLUSION RESTRICTION FOR NEVER TAKERS)

$$f(y(1) \mid U = 1, \mathbf{X} = \mathbf{x}, \theta_{11}^{UZ}) = f(y(0) \mid U = 1, \mathbf{X} = \mathbf{x}, \theta_{10}^{UZ}). \quad (3)$$

Assumption A2, which was made by Little and Yau (1998), and Yau and Little (2001) in their analyses of the JOBS-II data, states that the never taker distribution is the same under either assignment. This assumption is often reasonable, especially in blinded studies, and is sometimes necessary for identification. For these data, the assumption would be violated, for example, if never takers in the treatment group became more depressed as a result of their failure to take advantage of the intervention.

Under assumptions A1 and A2, the complier marginal distributions $f(y(1) \mid U \neq 1)$ and $f(y(0) \mid U \neq 1)$ are identified (Imbens and Rubin 1997). Additional assumptions are required to distinguish the two complier classes. Here we assume the complier marginal distributions are mixtures of distributions of known parametric form. In addition, we must also limit attention to cases where the mixtures are identified. For example, if the outcome is binary, a natural approach would be to model the complier distribution as a mixture of Bernoulli random variables. However, binomial mixtures are not identified (Titterington, Smith and Makov, 1985), precluding our approach in this case. On the other hand, continuous outcomes are often modeled as mixtures of normals; that representation is unique (up to a permutation of the components). See Titterington et al. (1985) for a general discussion of identification

of mixture distributions. We make the identification assumption explicit below.

A3 (MIXTURE IDENTIFIABILITY) The mixture densities are assumed to belong to a known parametric family. Further, for $Z = 0, D = 0$, $Z = 1, D = 0$, and $Z = 1, D = 1$, the parameters θ_{dz}^{DZ} of the component densities of the distributions $f(y \mid d, z, \mathbf{x}, \theta_{dz}^{DZ}, \lambda)$ are assumed to be identified.

A4 (ZERO EFFECT COMPLIER CLASS ASSUMPTION)

For U = 1,2,3, $\pi_u(\mathbf{x}, \lambda) > 0$. Further,

$$f(y(1) \mid U = 3, \mathbf{X} = \mathbf{x}, \theta_{31}^{UZ}) = f(y(0) \mid U = 3, \mathbf{X} = \mathbf{x}, \theta_{30}^{UZ}). \tag{4}$$

Assumption A4 states that all three class types have non-zero probability at each $\mathbf{x}$. For subjects in the zero effect class, $Y_i(0) = Y_i(1)$, implying the exclusion restriction (4).

Under assumptions A1 and A4, the likelihood (2) may be rewritten as:

$$L(\xi, \lambda) = \prod_{d,z} \prod_{i \in S_{d1z}^{DRZ}} \sum_{u \in A(d,z)} f(y_i(z_i) \mid u_i, \mathbf{x}_i, \theta_{uz}^{UZ}) \pi_u(\mathbf{x}_i, \lambda). \tag{5}$$

where $A(0,0) = \{1,2,3\}$, $A(0,1) = \{1\}$, $A(1,1) = \{2,3\}$ for the three class model.

For the two class model, assumption A4 is not needed, and may be replaced by the assumption $\pi_u(\mathbf{x}, \lambda) > 0$ for $U = 1, 2$; here $A(0,0) = \{1,2\}$, $A(0,1) = \{1\}$, $A(1,1) = \{2\}$.

3.2 *First Pass Reanalysis of the JOBS-II Data*

Table 1 reports means and standard deviations for variables used in Little and Yau's (1998) analyses of the 502 high risk subjects; their outcome is the difference between the depression score at 6 months and baseline, with negative scores indicating improvement. The mean difference in outcomes between treatment and control subjects of -.075 estimates the intent to treat estimand (ITT). 55 percent of the subjects are estimated to be compliers, leading to an instrumental variable estimate of -.136 for the CACE, similar to the mean difference between compliers and never takers in the treatment group. The balance between the covariates in

the treatment and control groups reflects the random assignment in the study. Never takers tend to be more assertive, younger, and of lower socio-economic status than compliers.

[Table 1 about here.]

Model M_{02} in Table 2, our baseline model, is the model Little and Yau (1998) estimated. They assumed the zero effect class $U = 3$ is empty, and for $Z = 0, 1$ and $U = 1, 2$, they modelled the outcome distributions $f(y(z) \mid u, \mathbf{x})$ as Normal with means $E(Y(z) \mid U = u, \mathbf{X} = \mathbf{x}) = \alpha_{uz}^{UZ} + \beta'\mathbf{x}$, with $\alpha_{10}^{UZ} = \alpha_{11}^{UZ}$ and common error variance $\sigma_{uz}^{2UZ} = \sigma^2$. The CACE is thus $\alpha_{21}^{UZ} - \alpha_{20}^{UZ}$. Of the previously described covariates, Little and Yau included only risk and baseline depression as regressors in the outcome models. They modeled compliance status using logistic regression: $\ln(\frac{\pi_2(\mathbf{x},\lambda)}{\pi_1(\mathbf{x},\lambda)}) = \lambda_0 + \lambda'\mathbf{x}$. They excluded the covariates used in the outcome models (risk and baseline depression) from the logistic regression.

In order to maintain comparability between our analyses and that of Little and Yau (1998), we also model the change in depression using the same covariates used in their analyses, using model M_{02} as a baseline for comparison with more general two and three class models.

[Table 2 about here.]

The Bayesian information criterion $-2\ln L + NP\ln n$ (hereafter BIC) is used as a model selection criterion, where L is the likelihood under the model, NP is the number of independent parameters estimated, and a lower BIC is indicative of a better model. Next, several less restrictive two class models were considered and compared with model M_{02} (BIC = 1545.89, NP = 14). In Model M_{12} (BIC = 1555.81 NP = 16) the error variances differ for never takers and compliers, and compliers have different error variances in the treatment and control groups. Model M_{22} (BIC = 1555.46, NP = 18) also relaxes the never-taker exclusion restriction $\alpha_{10}^{UZ} = \alpha_{11}^{UZ}$, $\sigma_{10}^{2UZ} = \sigma_{11}^{2UZ}$. Neither M_{12} nor M_{22} improves on model M_{02}, which is therefore the preferred two class model.

Analogous three class models were considered next. Here, compliance status was modelled using a multinomial logit model with never takers as the reference group: $\ln(\frac{\pi_u(\mathbf{x},\lambda)}{\pi_1(\mathbf{x},\lambda)}) = \lambda_{0u} + \lambda_u'\mathbf{x}$, $u = 2, 3$. Model M_{03} (BIC = 1561.31, NP = 23) is the three class model analogous to the baseline model M_{02}. Under this model, $\alpha_{10}^{UZ} = \alpha_{11}^{UZ}$, $\alpha_{30}^{UZ} = \alpha_{31}^{UZ}$, $\sigma_{uz}^{2UZ} = \sigma^2$ for $z = 0, 1$ and $u = 1, 2, 3$. In model M_{13} (BIC = 1541.59, NP = 26), which relaxes the equality constraint on the error variances, four error variances are estimated: $\sigma_{10}^{2UZ} = \sigma_{11}^{2UZ}$, σ_{20}^{2UZ}, σ_{21}^{2UZ}, $\sigma_{30}^{2UZ} = \sigma_{31}^{2UZ}$, Model M_{23} (BIC = 1549.48, NP = 28) also relaxes the never taker exclusion restriction $\alpha_{10}^{UZ} = \alpha_{11}^{UZ}$, $\sigma_{10}^{2UZ} = \sigma_{11}^{2UZ}$. This does not improve upon model M_{13}, which is the preferred three class model. Note that M_{13} is also preferred over the preferred two class model M_{02}.

Table 2 reports the estimated posterior probabilities $\hat{\pi}_u$ of class membership and CACE for Little and Yau's (1998) two class model M_{02} and the posterior probabilities and ECACE for the preferred three class model M_{13}. The never taker proportions are virtually identical in the two models, approximately 45 %. In model M_{13}, of the 54 % of the subjects estimated to be compliers, 44 % of these are in the effect class and 56 percent in the zero effect class. The estimated ECACE of -.389 is significantly different from 0 at the .05 level, and suggests more benefit for compliers in the effect class than the estimated CACE of -.310 in model M_{02}. Age and motivation are positively related to compliance in both models, with similar magnitude in the two effect classes in model M_{13}. Education and being single are positively associated with compliance in both models, apparently somewhat more so in the the zero class in model M_{13}, while assertiveness is negatively associated with compliance in both models, apparently more so in the effect class than in the zero effect class in model M_{13}. In model M_{02}, economic hardship is not significantly different from 0 at the .05 level, but in model M_{13}, hardship is negatively associated with compliance in the zero effect class. In both models, baseline depression is negatively associated with the change in depression. In

model M_{02}, the coefficient is -1.462, with a standard error of .173, suggesting that subjects became less depressed over the six month interval between the two measurements, whereas in model M_{13} the coefficient is -.904, with standard error .151, suggesting no change.

4. Compliance Mixture Modelling with Missing Data

In addition to the densities $f(y(z) \mid u, \mathbf{x}, \theta_{uz}^{UZ})$ for $u = 1, 2, 3$, $z = 0, 1$, and the class proportions $\pi_u(\mathbf{x}, \lambda)$, we now model the distribution of $R(z)$ given X and U: for $z = 0, 1$, r = 0,1, $\Pr(R(z) = r \mid u, \mathbf{x}, \psi_{uz}^{UZ})$. Let $\psi = (\psi_{10}'^{UZ}, ..., \psi_{31}'^{UZ})' = L\eta$, where η is the vector of distinct elements of ψ.

4.1 *Missing Data Assumptions*

To take into account missing outcome data, additional identifying assumptions are necessary. Yau and Little (2001) assumed the missing data were missing at random (MAR):

ASSUMPTION 5 (MISSING AT RANDOM)

$$Y \perp\!\!\!\perp R \mid Z, \mathbf{X}, U \tag{6}$$

for U observed,

$$Y, U \perp\!\!\!\perp R \mid Z, \mathbf{X} \tag{7}$$

for U unobserved.

For a two class model, U is observed if $Z = 1$, unobserved if Z = 0. For the three class model here, U is observed only if $Z = 1, D = 0$. Under MAR, the response process is "ignorable" if the parameters η governing the missing data mechanism are distinct from (ξ, λ).

For a randomized experiment with no always takers and a single complier class $U = 2$, Frangakis and Rubin (1999) pointed out that compliance status is observed in the treatment group, unobserved in the control group. The MAR assumption is unattractive here, implying

missingness depends on compliance status and covariates in the treatment group and only on covariates in the control group. As an alternative to assumption A5, they assumed the missing data are MAR given treatment assignment, observed covariates and compliance status U. We extend this assumption here:

A6 (EXTENDED LATENT IGNORABILITY)

$$Y \perp\!\!\!\perp R \mid U, Z, \mathbf{X}. \tag{8}$$

For the two class case with no covariates, assuming randomization and the never taker exclusion restriction, Frangakis and Rubin (1999) then construct a methods of moments estimator of the ITT, replacing the stronger MAR assumption with the weaker latent ignorability assumption. However, this also necessitates an additional identifying assumption. To that end, they proposed the never taker response exclusion restriction $R_i(0) = R_i(1)$ if $U_i = 1$. It is worth noting that the combination of assumption A6 with the never taker response exclusion restriction is neither weaker nor stronger than the MAR assumption A5.

In their two class model applied to a randomized study of breast self examination, Mealli et al. (2004) instead combine the latent ignorability assumption A6 with the complier response exclusion restriction $R_i(0) = R_i(1)$ if $U_i = 2$. As above, this combination of assumptions is neither stronger nor weaker than the MAR assumption. Similarly to us, Mealli et al. (2004) use a parametric mixture model; however, in their analysis, the outcome Y is binary.

In our analysis, based on normal mixtures, both two and three class models are identified under the MAR assumption A5. Under assumption A6, response exclusion restrictions of the form above are not needed, but in our three class models alternative identifying assumptions about the response process are needed to distinguish the complier classes. To that end, we consider two possibilities. If the response process does not depend on the effectiveness of the intervention, it would be reasonable to expect the two types of compliers to behave in the same fashion, leading to assumption A7 below. However, if responding is positively related

to the effectiveness of the intervention, zero effect compliers might be expected to behave like never takers, leading to assumption A8 below.

A7 (ZERO EFFECT/EFFECT CLASS HOMOGENEITY)

For u = 2,3 and z = 0,1,

$$\Pr(R(z) = 1 \mid U = 2, \mathbf{X} = \mathbf{x}) = \Pr(R(z) = 1 \mid U = 3, \mathbf{X} = \mathbf{x}). \tag{9}$$

Assumption A7 states that, conditional on covariates and treatment assignment, compliers in the effect and zero effect classes are equally likely to report the outcome Y. Note that assumptions A6 and A7 are jointly weaker than the MAR assumption A5.

A8 (NEVER TAKER/ZERO EFFECT CLASS HOMOGENEITY)

For u = 1,3 and z = 0,1,

$$\Pr(R(z) = 1 \mid U = 1, \mathbf{X} = \mathbf{x}) = \Pr(R(z) = 1 \mid U = 3, \mathbf{X} = \mathbf{x}). \tag{10}$$

Assumption A8 states that, conditional on covariates and treatment assignment, never takers and zero effect compliers are equally likely to report the outcome Y. Assumptions A6 and A8 are jointly weaker than the MAR assumption A5.

Under assumptions A1 and A4 (as in section 3) and either the missing data assumption A5 or A6, the likelihood to be maximized with respect to the parameters may be written as:

$$\begin{aligned} L(\xi, \lambda, \eta) = \prod_{d,z} \{ & \prod_{i \in S^{DRZ}_{d1z}} \sum_{u \in A(d,z)} f(y_i(z_i) \mid u, \mathbf{x}_i, \theta^{UZ}_{uz}) \Pr(R_i(z_i) = 1 \mid u, \mathbf{x}_i, \psi^{UZ}_{uz}) \pi_u(\mathbf{x}_i, \lambda) \\ & \prod_{i \in S^{DRZ}_{d0z}} \sum_{u \in A(d,z)} \Pr(R_i(z_i) = 0 \mid u, \mathbf{x}_i, \psi^{UZ}_{uz}) \pi_u(\mathbf{x}_i, \lambda) \}. \end{aligned} \tag{11}$$

When assumption A6 is combined with either assumption A7 or A8, further restrictions on $\Pr(R_i(z_i) = 1 \mid u, \mathbf{x}_i, \psi^{UZ}_{uz})$ are imposed in (11).

4.2 *Second Pass Reanalysis of the JOBS-II Data*

We use the 657 high risk subjects with complete covariate data to illustrate several two and three class models for these data. Previously, to maintain comparability with the analysis

in Little and Yau (1998), the outcome was the difference between depression at six months and baseline depression. In this section, we model depression at 6 months, with baseline depression now included as a covariate; further, as all subjects are high risk, the risk variable itself is no longer included as a covariate. Attention herein focuses on estimating the class proportions and ECACE and the sensitivity of these estimates to assumptions about the response process and the number of classes.

[Table 3 about here.]

Table 3 presents summary results from the models of primary interest that were fitted using maximum likelihood estimation. As in the previous section, the regression coefficients for the 8 covariates are constrained to be equal across all classes, in both treatment and control groups; in addition, in the 2 class models, we estimate 1 intercept and 1 residual variance for the never takers, 2 intercepts and 2 residual variances for the compliers, and in the 3 class models an additional intercept and residual variance for the 0 effect compliers. 9 parameters are used to predict class membership in the 2 class models, 18 parameters in the 3 class models. Coefficients for the 8 covariates in the response process are constrained to be equal across all classes, and in both treatment and control groups. The remaining parameters are for the intercepts in the logistic models of the response process; the number of distinct intercepts depends on the missing data assumptions.

Model M_{2MAR}, in which the missing data are assumed to be MAR, is the 2 class analogue to model M_{12} of the previous section, while model M_{2LI} is the analogous 2 class model where the missing data are assumed latently ignorable. In both models, the complier posterior probability is estimated as .53, similar to the estimates in the previous section. Both models suggest greater benefit for the intervention than that suggested by the 2 class model M_{02} of the previous section. Model M_{2LI} was also fit a) with the never taker response exclusion re-

striction imposed and b) with the complier response exclusion restriction imposed, obtaining in both cases results very similar to those obtained with model M_{2MAR}.

Model M_{3MAR} is the baseline 3 class model in which the missing data are assumed MAR: $\Pr(R(0) = 1 \mid U = 1, \mathbf{x}) = \Pr(R(0) = 1 \mid U = 2, \mathbf{x}) = \Pr(R(0) = 1 \mid U = 3, \mathbf{x})$, $\Pr(R(1) = 1 \mid U = 2, \mathbf{x}) = \Pr(R(1) = 1 \mid U = 3, \mathbf{x})$. The complier posterior probability is estimated as .54, similar to the two class missing data models, with about 37 % of the compliers in the zero effect class. Interestingly, even with a substantial zero effect class, the estimated ECACE of -.240 suggests less benefit from the intervention than that suggested by the two class models. Similarly, this estimate is smaller than the estimated ECACE of -.389 from the preferred 3 class model M_{13} of the previous section.

Model M_{3ELI7} imposes the extended latent ignorability assumption A6 and the zero effect/effect class homogeneity assumption A7. The estimated posterior probabilities and ECACE are virtually identical to those for model M_{3MAR}, which is preferred on the basis of BIC. Model M_{3ELI8} imposes the extended latent ignorability assumption A6 and the never taker/zero effect class homogeneity assumption A8. While the estimated posterior probabilities are similar to those in the other three class models considered, here the estimated ECACE of -.653 is considerably larger in magnitude, with a one tailed p value of .046.

Further inspection of the BIC values in table 3 suggest that M_{2LI}, the two class model with latently ignorable missingness, receives the least support, while the 3 class MAR model M_{3MAR} would be chosen as the preferred model for these data. However, the BIC values do not suggest strong support for M_{3MAR} over the 2 class MAR model or the other 3 class models. Using the information from the various models suggests the intervention lowers depression at six months between about .03 points (using the lower bound from the 95 % confidence interval for ECACE from M_{3MAR}) to about 1.44 points (using the upper bound from the 95 % confidence interval for ECACE from M_{3ELI8}.

5. Discussion

In randomized studies with non-compliance, subjects are typically split into classes by compliance type, with interest centering on estimation of the average effect in the complier class. This paper proposes further subdividing compliers into a zero effect class, where the intervention does not affect the outcome, and an effect class, where the intervention affects the outcome. For the Jobs-II study, almost half the subjects are never takers, approximately 30 percent are effect class compliers, and approximately 20 percent are in the zero effect complier class. The large proportion of never takers could reflect a) problems in the program delivery, as would be the case if subjects found attending multiple sessions onerous, and/or b) the program content, as would be the case if many never takers felt they already knew how to look for jobs, and/or c) neither, to the extent that some of the never takers would not participate in any intervention. Further study would be needed before any recommendations to deal with this situation could be made. A substantial proportion of the compliers (approximately 40 percent) belong to the zero effect class, and the analyses do not suggest that the ECACE is especially large in the effect class. In this study, where it is reasonable to think that the compliers improved their search skills, the analyses suggest that improved skills do not lead to less depression in the zero effect class and only lead to a modest decrease in the effect class. This suggests replacing or augmenting the intervention with components designed to target other pathways that could lead to a decrease in depression, for example, teaching subjects how to recognize depression and seek attention for depression using available social services.

As in Little and Yau (1998) and other analyses of these data (for example, Jo 2002), our analysis assumes that the mixture components are normal, conditional on covariates. To assess whether or not a zero effect class is needed, we compared models with and without a zero class, using the Bayesian information criterion (BIC). With no covariates, the model

fit might be assessed by comparing the empirical cumulative distribution function with that estimated under the model (Aitkin 1997), but with many covariates, as here, this would not be feasible. One might, however, consider using a robust alternative, such as a mixture of split-t distibutions, as in Li, Villani and Kohn (2010). A number of other extensions of the framework herein would also be useful. In many cases, especially those involving a more complicated treatment regimen, the majority of subjects assigned to treatment do not follow the regimen exactly, but follow it to some degree. Consideration of studies where control group subjects can access the treatment and studies with multiple arms would also be useful. In addition, we have treated the case of a one time intervention with a single outcome measure measured at a single point in time. Our treatment of missing data assumes the missing outcome data are either missing at random or latently ignorable. It would also be useful to consider the case where missingness depends on the outcome (Chen, Geng and Zhou 2009). Generalization to multiple and repeated outcomes with missing data would also be desirable (see for example Beunckens, Molenberghs, Verbeke and Mallinckrodt (2008), Muthén, Asparohouv, Hunter and Leuchter (2011), Yuan and Little (2009). Recent work on longitudinal causal inference has paid much attention to the case of sequentially randomized experiments, in which a subject's treatment over time evolves according to his/her previous history and treatments. Consideration of this case would also be useful.

ACKNOWLEDGEMENTS

We thank the associate editor and two anonymous reviewers for helpful comments on a previous draft of this paper.

References

Aitkin, M. (1997). Discussion of "On Bayesian Analysis of Mixtures with an Unknown Number of Components" by S. Richardson and P. J. Green. *Journal of the Royal*

Statistical Society, Series B **59**, 764-768.

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**, 444-472.

Beunckens, C., Molenberghs, G., Verbeke, G. and Mallinckrodt, C. (2008). A Latent-Class Mixture Model for Incomplete Longitudinal Gaussian Data. *Biometrics* **64**, 96-105.

Bloom, H. S. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review* **8**, 225-246.

Derogatis, L.R., Lipman, R.S.,Rickles,K.,Uhlenhuth,E.H. and Covi,L. (1974). The Hopkins Symptom Checklist (HSCL). in *Psychological Measurements in Psychopharmacology: Modern Problems in Pharmacopsychiatry* (Vol. 7), ed. P. Pichot, New York: Karger, pp. 79-110.

Cheng, J., Qin, J., and Zhang, B. (2009). Semiparametric Estimation and Inference for Distributional and General Treatment Effects. it Journal of the Royal Statistical Society, Series B **71**,881-904.

Frangakis, C.E. and Rubin, D.B. (1999). Adressing Complications of Intent to Treat Analysis in the 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.

Imbens, G.W. and Rubin,D.B. (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies* **65**, 555-574.

Jin, H. and Rubin, D.B. (2008). Principal Stratification for Causal Inference with Extended Partial Compliance. *Journal of the American Statistical Association* **103**, 101-111.

Jo, B. (2002). Estimating Intervention Effects with NonCompliance: Alternative Model Specifications. *Journal of Educational and Behavioral Statistics* **27**, 385-430.

Kowatch, R.A., Suppes, T., Carmody, T.J., Bucci, J.P., Hume, J.H., Kromelis, M., Emslie, G.J., Weinberg, W.A., and Rush, J. (2002). Effect Size of Lithium, Divalproex Sodium, and Carbamazepine in Children and Adolescents with Bipolar Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 713-720.

Li, F., Villani, M., and Kohn, R. (2010). Flexible Modeling of Conditional Distributions Using Smooth Mixtures of Asymmetric Student t Densities. *Journal of Statistical Planning and Inference* **140**, 3638-3654.

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* **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**, 202-227.

Muthén, B. and Shedden, K. (1999). Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics* **66**, 463-469.

Muthén, B. and Asparouhov, T. (2009). Growth Mixture Modeling: Analysis with non-Gaussian Random Effects. in *Longitudinal Data Analysis* , eds. G. Fitzmaurice, M. Davidian, G. Verbeke, G. Molenberghs, Boca Raton: Chapman and Hall/CRC Press, pp. 143-165.

Muthén, B., Asparouhov, T., Hunter, A. and Leuchter, A. (2011). Growth Modeling with Non-Ignorable Dropout: Alternative Analyses of the Star*D Antidepressant Trial. *Psychological Methods* **16**, 17-33.

Muthén, B. and Muthén, L. (1998-2011). *Mplus User's Guide*, 6th edn. Los Angeles, CA: Muthén and Muthén.

Peng, Y., Little, R.J. 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.

Price, R.H., van Ryn, M. and Vinokur, A.D. (1992). Impact of Preventive Job Search Intervention on the Likelihood of Depression Among the Unemployed. *Journal of Health and Social Behavior* **33**, 158-167.

Small, D., Ten Have, T., Joffe, M. and Cheng, J. (2006). Random Effects Logistic Models for Analysing Efficacy of a Longitudinal Randomized Treatment with Non-Adherence. *Statistics in Medicine* **25**, 1981-2007.

Titterington, D.M., Smith, A.F.M and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

Vinokur, A.D., Price, R.H. and Shul, Y. (1995). Impact of JOBS Intervention on Unemployed Workers Varying in Risk for Depression. *American Journal of Community Psychology* **23**, 39-74.

Walker, S. A., White, I.R. and Babikar,A.G. (2004). Parametric Randomization-Based Methods for Correcting for Treatment Changes in the Assessment of the Causal Effect of Treatment. *Statistics in Medicine* **23**, 571-590.

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.

Yuan, Y. and Little, R.J (2009). Mixed-Effect Hybrid Models for Longitudinal Data with Nonignorable Dropout. *Biometrics* **65**, 478-486.

Table 1

Means and standard deviations (in parentheses) for 502 high risk subjects in the Jobs-II study.

Variable	Z = 0		Z = 1		Z = 1,D = 0		Z = 1,D=1	
Outcome	-.383	(.803)	-.458	(.763)	-.390	(.784)	-.513	(.740)
Age	36.17	(9.75)	36.79	(10.06)	33.31	(9.57)	39.68	(9.56)
Education	13.34	(1.98)	13.38	(2.05)	12.89	(1.90)	13.79	(2.08)
Marital status	.58	(.49)	.64	(.48)	.63	(.48)	.65	(.48)
Economic hardship	3.47	(.95)	3.68	(.82)	3.79	(.84)	3.60	(.80)
Race	.18	(.39)	.20	(.40)	.25	(.43)	.15	(.36)
Depression	2.49	(.29)	2.43	(.30)	2.44	(.31)	2.42	(.30)
Motivation	5.32	(.80)	5.34	(.82)	5.14	(.81)	5.50	(.79)
Assertiveness	3.03	(.88)	3.09	(.93)	3.24	(.92)	2.96	(.92)
Risk	1.69	(.19)	1.67	(.22)	1.68	(.21)	1.67	(.21)
Sample size	167		335		152		183	

Table 2

Estimates of ECACE, class membership and parameters of compliance models for 502 high risk subjects.

	Model M_{02}		Model M_{13}			
BIC and NP	1545.89	14	1541.49	26		
ECACE	-.310	(.130)	-.389	(.187)		
$\hat{\pi}_1$	.458		.460			
$\hat{\pi}_2$	.542		.236			
$\hat{\pi}_3$			.304			
VARIABLE	λ		$\lambda_{\mathbf{2}}$		$\lambda_{\mathbf{3}}$	
Age	.079	(.015)	.080	(.018)	.077	(.018)
Education	.300	(.068)	.205	(.091)	.363	(.081)
Marital Status	.540	(.283)	.371	(.345)	.664	(.335)
Economic Hardship	.159	(.152)	.253	(.205)	-.433	(.196)
Race	-.499	(.317)	-.666	(.476)	-.423	(.360)
Motivation	.667	(.157)	.657	(.210)	.652	(.188)
Assertiveness	-376	(.143)	-.540	(.201)	-.138	(.180)

Table 3

Estimates of ECACE and class membership for some models with different missing data assumptions for 657 high risk subjects.

Model	BIC	NP	ECACE		Class Proportions		
M_{2MAR}: 2 class MAR	2468.69	34	-.382	(.120)	.471	.529	
M_{2LI}: 2 class LI	2474.67	35	-.404	(.139)	.471	.529	
M_{3MAR}: 3 class MAR	2461.66	45	-.240	(.107)	.462	.340	.198
M_{3ELI7}: 3 class ELI + A7	2468.11	46	-.239	(.106)	.462	.340	.198
M_{3ELI8}: 3 class ELI + A8	2467.00	46	-.653	(.393)	.472	.304	. 224