Sensitivity of Causal Effects Under Ignorable and Latent Ignorable Missing-Data Mechanisms

Booil Jo*,†

Division of Social Research Methodology, Graduate School of Education & Information Studies, University of California, Los Angeles

(Rough Draft: December 24, 2002)

* The study was supported by National Institute on Drug Abuse and National Institute of Mental Health Grant MH40859 (Hendricks Brown and Bengt Muthén, Co-Principal Investigators). I thank participants in Hendricks Brown's Prevention Science Methodology Group and Bengt Muthén's Research Apprenticeship Course for helpful feedback. Correspondence to: Booil Jo, Division of Social Research Methodology, Graduate School of Education & Information Studies, University of California, Los Angeles, CA 90095-1521.

†E-mail: booil@ucla.edu

1 INTRODUCTION

Randomized trials involving human participants often suffer from both noncompliance and missingness in outcomes (nonresponse or dropout). In these situations, it seems quite natural to link individuals' compliance type to their dropout behavior or response to surveys. For example, treatment noncompliers may be less likely to respond to follow-up surveys or more likely to drop out from the study than compliers. These settings lead to potentially nonignorable missing-data mechanisms, given that missingness is related to latent compliance class, which is partly unobserved.

If the probability of missingness in outcomes depends only on observed data, the missing data are considered as missing at random (Little & Rubin, 1987). Under the assumption of missing at random (MAR), the missing-data mechanism is ignorable for likelihood-based inferences. In other words, whether the mechanism leading to missing data is explicitly specified in the model does not influence inferences. However, if missingness is attributable to unobserved data, the missing-data mechanism is nonignorable. In this case, likelihood-based inferences can be sensitive to whether and how the missing-data mechanism is specified in the statistical model.

Frangakis and Rubin (1999) suggested their novel idea of modeling missing-data mechanisms related to compliance behavior of study participants. They defined "latent ignorability", where potential outcomes and associated potential nonresponse indicators are independent within each level of the latent compliance strata. Since compliance status is not completely observed for every individual, the missing-data mechanism related to compliance is no longer ignorable. This latent ignorability assumption conditional on compliance is weaker than the conventional ignorability

(MAR) assumption and is potentially useful in investigating the sensitivity of the ITT effect and the CACE estimates to violation of ignorability.

Based on simulation studies, Frangakis and Rubin (1999) demonstrated that models assuming latent ignorability generally perform better than (or at least as well as) models assuming ignorability even when the latent ignorability conditions are not satisfied. The substantial advantage of assuming latent ignorability found in Frangakis and Rubin (1999) was, however, based only on conditions where the response rate is quite different for compliers and noncompliers (80% vs. 50%: big deviation from ignorability). Since Frangakis and Rubin (1999), a systematic investigation has not been conducted to define conditions under which models assuming ignorability or latent ignorability are most sensitive, and to define conditions under which substantial differences can be expected between the two missing data approaches. Further, it is not well known whether assuming latent ignorability instead of ignorability is always advantageous.

Compound exclusion (Frangakis & Rubin, 1999) is one of the key structural assumptions in identifying principal effects (Frangakis & Rubin, 2002) under latent ignorability. Under this assumption, the difference in outcomes and response rates between the treatment and the control condition is allowed for compliers, but is not allowed for never-takers (individuals who would not receive the treatment regardless of whether it is offered) or for always-takers (individuals who would receive the treatment regardless of whether it is offered). The plausibility of the standard exclusion restriction in observed outcomes and the impact of violating the assumption have been previously discussed (Angrist et al., 1996; Frangakis et al., 2002; Hirano

et al., 2000; Imbens & Rubin, 1997; Jo, in press, 2002; Shadish, Cook, & Campbell, 2002; West & Sagarin, 2000). When the exclusion restriction is violated in observed outcomes, the causal effect of treatment not only can be understated, but also can be exaggerated depending on how the assignment of treatment affects noncompliers. However, little is known about the sensitivity of causal effect estimates to violation of the exclusion restriction in outcome missingness. Since the violation of the exclusion restriction in observed outcomes affects both models assuming ignorability and latent ignorability, this study will focus its investigation on the violation of the exclusion restriction in outcome missingness, which affects only latent ignorability models.

The plausibility of compound exclusion restriction is often hard to predict in practice given various possibilities and competing theories. Besides, the directions and magnitudes of deviation from the exclusion restriction can be different for observed outcomes and missingness of the outcomes. In examining the degree of deviation from ignorability, observed or estimated response rates without considering possible violation of compound exclusion could be misleading. For example, when there are only two compliance classes (e.g., compliers and never-takers), observed data in the treatment condition may indicate a clear difference in response rates between the two compliance classes. However, the same trend may not hold in the absence of treatment. Or, estimated response rates assuming compound exclusion may show a big difference between the two compliance classes in the control condition, but this finding may not hold if compound exclusion is violated. To gauge the relative benefit of models assuming ignorability and latent ignorability given these uncertainties, one needs to carefully consider reasons why (or why not) compliers and noncompliers

would show different response rates when assigned to the control condition (deviation from ignorability), and why (or why not) noncompliers would show any effects of treatment assignment on their response rates (deviation from compound exclusion). The answers to these questions could be quite obvious in some situations, but not in other situations.

This study focuses on the fact that identification of causal effects under latent ignorability relies on the generalized (compound) exclusion restriction (i.e., both on the outcomes and missingness of outcomes), whereas identification of causal effects under ignorability relies on the standard exclusion restriction (i.e., only on the outcomes). The assumption of latent ignorability is weaker than ignorability in the sense that it allows interaction between outcome missingness and partially observed compliance class, but stronger in the sense that the exclusion restriction is additionally imposed on outcome missingness. In other words, the relative benefit of models assuming latent ignorability and standard ignorability depends on degrees of deviation from compound exclusion and ignorability. On the basis of Monte Carlo Simulations, this study demonstrates sensitivity of causal effect estimates under ignorability and latent ignorability to various degrees of deviation from the standard ignorability and compound exclusion. It is also demonstrated that sensitivity of causal effect estimates to model misspecification can be somewhat reduced in the presence of covariates that are good predictors of compliance. Maximum likelihood estimation using the EM algorithm is employed in the estimation of causal effects in the study.

2 UNDERLYING ASSUMPTIONS

Assume the simplest experimental setting where there is only one outcome measure (Y), its missing indicator is R (1 = responded, 0 = not responded), treatment assignment (Z) is binary (1 = treatment, 0 = control), and the treatment received (D) has only two levels (1 = received, 0 = not received). The behavior types (C_i) of the subjects based on combinations of Z and D can be classified into four categories based on Rubin's causal model approach, where the possibility of statistical causal inference is built at the individual level (Holland, 1986; Rubin, 1978, 1980). Angrist et al. (1996) labeled the four categories as complier, never-taker, defier, and always-taker. Let $D_i(1)$ denote the potential treatment receipt status for individual i when assigned to the treatment condition, and $D_i(0)$ denote the potential treatment receipt status for individual i when assigned to the control condition. Compliers are subjects who do what they are assigned to do ($D_i(1) = 1$ and $D_i(0) = 0$). Never-takers are subjects who do not receive the treatment even if they are assigned to the treatment condition ($D_i(1) = 0$ and $D_i(0) = 0$). Defiers are the subjects who do the opposite of what they are assigned to do ($D_i(1) = 0$ and $D_i(0) = 1$). Always-takers are the subjects who always receive the treatment, no matter which condition they are assigned to ($D_i(1) = 1$ and $D_i(0) = 1$). Let $Y_i(1, D_i(1))$ and $R_i(1, D_i(1))$ denote the potential outcome and its missing indicator for individual i with treatment receipt status D_i when $Z_i = 1$, and $Y_i(0, D_i(0))$ and $R_i(0, D_i(0))$ denote the potential outcome and its missing indicator for individual i with treatment receipt status D_i when $Z_i = 0$.

Among these four types of subjects, the emphasis is usually given to the estimation of causal effect of treatment assignment for compliers (e.g., Angrist, Imbens &

Rubin, 1996; Bloom, 1984; Hirano, Imbens, Rubin, & Zhou, 2000; Imbens & Rubin, 1997; Jo, in press; Little & Yau, 1998). The following assumptions are critical in the identification of CACE.

– Randomization: Treatment assignment is random.

– Stable Unit Treatment Value (SUTVA: Rubin, 1978, 1980, 1990): Potential outcomes for each person are unrelated to the treatment status of other individuals.

– Monotonicity (Imbens & Angrist, 1994): There are no defiers.

– Exclusion Restriction (Angrist et al., 1996): For never-takers and always-takers, the distributions of the potential outcomes are independent of the treatment assignment. That is, $Y_i(0, D_i(0)) = Y_i(1, D_i(1))$ for units with $D_i(0) = D_i(1) = 0$ or $D_i(0) = D_i(1) = 1$.

– Compound Exclusion (Frangakis & Rubin, 1999): For never-takers and always-takers, the distributions of the potential outcomes and associated potential nonresponse indicators are independent of the treatment assignment. That is, $Y_i(0, D_i(0)) = Y_i(1, D_i(1))$ and $R_i(0, D_i(0)) = R_i(1, D_i(1))$ for units with $D_i(0) = D_i(1) = 0$ or $D_i(0) = D_i(1) = 1$.

When ignorability is assumed in the estimation, the standard exclusion restriction is sufficient. When latent ignorability is assumed in the estimation, compound exclusion is necessary to obtain unique ML estimates.

3 Causal Effect Estimation

For simplicity, let us assume that there are only two possible compliance behavior types (C_i). That is,

$$C_i = \begin{cases} c\,(\text{complier}) & \text{if } D_i(1) = 1, \text{ and } D_i(0) = 0 \\ n\,(\text{never-taker}) & \text{if } D_i(1) = 0, \text{ and } D_i(0) = 0, \end{cases}$$

where $C(t) = \{i \mid C_i = t\}$ for $t \in \{c, n\}$.

Let $c_i = 0$ and $n_i = 1$ if $i \in C(n)$, and $c_i = 1$ and $n_i = 0$ if $i \in C(c)$. A continuous outcome Y for individual i is

$$Y_i \;=\; \alpha_n\, n_i + \alpha_c\, c_i + \gamma_n\, n_i\, Z_i + \gamma_c\, c_i\, Z_i + \lambda_n\, n_i\, X_i + \lambda_c\, c_i\, X_i + \epsilon_{in}\, n_i + \epsilon_{ic}\, c_i, \quad (1)$$

where where α_c and α_n are intercepts for compliers and never-takers. γ_c is the average causal effect of treatment assignment for compliers (CACE: complier average causal effect). γ_n is the average causal effects of treatment assignment for never-takers. λ_c and λ_n are covariate effects on the outcome. ϵ_{ci}, and ϵ_{ni} are normally distributed residuals of compliers and never-takers.

The logistic regression of C on X is described as

$$logit(\pi_{ci}) \;=\; \beta_0 + \beta_1\, X_i, \quad (2)$$

where π_{ci} is the probability of being a complier, π_{ni} $(1 - \pi_{ci})$ is the probability of being a never-taker, β_0 is a logit intercept, and β_1 is a vector of logit coefficients.

The logistic regression of missing (response) indicator R (1 = responded, 0 = not responded) on C and Z is described as

$$logit(\pi_{Ri}) \;=\; \Lambda_{0n}\, n_i + \Lambda_{0c}\, c_i + \Gamma_n\, n_i\, Z_i + \Gamma_c\, c_i\, Z_i + \Lambda_{1n}\, n_i\, X_i + \Lambda_{1c}\, c_i\, X_i, \quad (3)$$

where Λ_{0n} and Λ_{0c} are logit intercepts of noncompliers and compliers. Γ_n represents the effect of treatment assignment on R for noncompliers, and Γ_c represents the effect of treatment assignment on R for compliers. Λ_{1n} and Λ_{1c} represent covariate effects on R for noncompliers and compliers. The proportion of responders π_R may vary across compliers (π_{Rc}) and never-takers (π_{Rn}).

This study employs a maximum likelihood estimation approach in the estimation of CACE. The unknown compliance status (C) in the control condition is handled as missing data via the EM algorithm (Dempster, Laird, & Rubin, 1977; Little & Rubin, 1987; McLachlan & Krishnan, 1997; Tanner, 1996). Parametric standard errors are computed from the information matrix of the ML estimator using both the first- and the second-order derivatives under the assumption of normally distributed outcomes. ML-EM estimation of CACE assuming ignorability and latent ignorability was carried out by the M*plus* program (Muthén & Muthén, 1998-2001). The general framework of ML-EM estimation under latent ignorability considering both categorical and continuous latent variables are outlined in Muthén and Brown (2001).

Viewing C as partly missing data, the complete-data log likelihood can be expressed for the general model as

$$\sum_{i=1}^{n}(log[C_i \mid X_i] + log[Y_i \mid C_i, X_i]), \tag{4}$$

where X includes a vector of covariates and treatment assignment Z. With missing data, the Y part of Equation (4) expands to $[Y_i^{obs}, Y_i^{miss}, R_i \mid C_i, X_i]$. We are interested in

$$[Y_i^{obs}, R_i \mid C_i, X_i] = \int [Y_i^{obs}, Y_i^{mis} \mid C_i, X_i]\, [R_i \mid Y_i^{obs}, Y_i^{mis}, C_i,\, X_i]\; dY_i^{mis}. \tag{5}$$

If ignorability is assumed, Equation (5) simplifies to

$$[Y_i^{obs}, R_i \mid C_i, X_i] = [Y_i^{obs} \mid C_i, X_i]\,[R_i \mid Y_i^{obs},\, X_i], \tag{6}$$

where the two terms of Equation (6) do not share parameters so that the last term can be ignored and maximization can focus on the $[Y_i^{obs} \mid C_i, X_i]$.

If latent ignorability is assumed, the $[R_i \mid C_i, *]$ term cannot be ignored, but would need to be updated in the EM algorithm to take into account the updated E step information on C_i. In this way, ignorability does not hold for the mixture distributions of $[Y \mid X]$. Therefore, models assuming latent ignorability will focus on

$$[Y_i^{obs} \mid C_i, X_i]\,[R_i \mid C_i,\, X_i]. \tag{7}$$

4 SIMULATION STUDY

This section demonstrates the quality of the CACE estimate in varying conditions of standard ignorability (SI), latent ignorability (LI), and compound exclusion (CE). For simplicity, the simulation study considers only compliers and never-takers. The simulation results presented in this section are based on 500 replications with a sample size of 2000. Equal probability of treatment/control assignment and 50% compliance rate were used as true values for all simulation settings in this study. The outcomes Y_i are normally distributed with variance of one. The outcome mean is 2.5 for compliers and 1.0 for never-takers when $Z = 0$. The true average causal effect of treatment assignment on the outcome for compliers (CACE) is 0.5, which corresponds to 0.5 in terms of effect size. The true average causal effect on the outcome for never-takers is fixed at zero in every simulation setting. The average response rate is 50% when $Z = 0$. The effect of deviation from SI is examined in situations, where 1) response rate is 50% for both compliers and never-takers when $Z = 0$. (i.e., SI holds), 2)

response rate is 55% for never-takers and 45% for never-takers (10% difference), 3) response rate is 60% for never-takers and 40% for never-takers (20% difference), and 4) response rate is 65% for never-takers and 35% for never-takers (30% difference). The true average causal effect of treatment assignment on outcome missingness for compliers is 2.5 in terms of odds ratio. The effect of deviation from CE is examined in situations, where 1) the average causal effect of treatment assignment on outcome missingness for never-takers is 1.0 in terms of odds ratio (i.e., CE holds), 2) the odds ratio is 1.5, 3) the odds ratio is 2.0, and 4) the odds ratio is 2.5. Covariates are not included in simulations shown in Tables 1.a and 1.b. One covariate is included in simulations shown in Tables 2.a and 2.b.

Table 1.a. Average CACE estimates when assuming standard ignorability for missingness of outcomes (SE and coverage in parentheses).

Deviation from CE	Deviation from SI (difference between π_{R_n} and π_{R_c})			
	0%	10%	20%	30%
OR (Γ_n) = 1.0	0.502 (0.104, 95.8%)	0.429 (0.102, 88.6%)	0.369 (0.100, 74.4%)	0.319 (0.098, 54.0%)
OR (Γ_n) = 1.5	0.502 (0.104, 95.2%)	0.430 (0.101, 89.2%)	0.369 (0.099, 74.4%)	0.318 (0.098, 54.0%)
OR (Γ_n) = 2.0	0.502 (0.104, 95.4%)	0.430 (0.101, 89.2%)	0.369 (0.099, 73.4%)	0.319 (0.097, 52.8%)
OR (Γ_n) = 2.5	0.502 (0.103, 95.4%)	0.430 (0.101, 88.6%)	0.369 (0.099, 73.4%)	0.319 (0.097, 52.4%)

Table 1.a shows the sensitivity of the CACE estimate to violation of CE and SI when SI and the standard exclusion restriction are assumed for the missing-data mechanism. Average standard errors and 95% confidence interval coverage probabilities are shown in the parentheses. It is shown that CACE estimates are sensitive to

deviation from SI, but not to deviation from CE. If compliers and never-takers show a large difference in response rates, CACE estimates can be seriously biased when assuming SI.

Table 1.b. Average CACE estimates when assuming latent ignorability for missingness of outcomes (SE and coverage in parentheses).

Deviation from CE	Deviation from SI (difference between π_{R_n} and π_{R_c})			
	0%	10%	20%	30%
OR (Γ_n) = 1.0	0.502 (0.111, 95.4%)	0.496 (0.104, 94.8%)	0.495 (0.099, 94.2%)	0.497 (0.092, 95.0%)
OR (Γ_n) = 1.5	0.384 (0.118, 84.8%)	0.384 (0.111, 81.6%)	0.388 (0.104, 81.8%)	0.396 (0.097, 83.2%)
OR (Γ_n) = 2.0	0.309 (0.125, 68.2%)	0.312 (0.116, 64.2%)	0.319 (0.108, 62.8%)	0.331 (0.102, 60.2%)
OR (Γ_n) = 2.5	0.258 (0.130, 53.2%)	0.261 (0.121, 49.2%)	0.270 (0.112, 47.8%)	0.282 (0.105, 44.2%)

Table 1.b shows the sensitivity of the CACE estimate to violation of CE and SI when LI and compound exclusion is assumed for the missing-data mechanism. It is shown that CACE estimates are sensitive to deviation from CE, but not to deviation from SI. If never-takers show a large effect of treatment assignment on response rates, CACE estimates can be seriously biased when assuming LI. The simulation results shown in Tables 1.a and 1.b indicate that the relative benefit of assuming SI and LI depends on which assumption (SI or CE) is more plausible or more severely biased.

The simulation results in Tables 2.a and 2.b show whether the information from covariates associated with compliance behavior increases insensitivity of CACE estimates to deviation from SI or CE in models assuming SI and LI. For simplicity, one continuous covariate ($X_i \sim N(0,1)$) that only predicts C (odds ratio = 2.0) is

used. It is demonstrated that including a covariate resulted in a noticeable, but not dramatic improvement in the quality of CACE estimates. The results show that covariate information reduces sensitivity of CACE estimates to the violation of SI in models assuming SI (Table 2.a), and reduces sensitivity of CACE estimates to the violation of CE in models assuming LI (Table 2.b).

Table 2.a. Average CACE estimates when assuming standard ignorability for missingness of outcomes in the presence of a covariate (SE and coverage in parentheses).

Deviation from CE	Deviation from SI (difference between π_{R_n} and π_{R_c})			
	0%	10%	20%	30%
OR (Γ_n) = 1.0	0.506 (0.102, 94.0%)	0.448 (0.100, 92.2%)	0.394 (0.097, 81.8%)	0.350 (0.096, 62.6%)
OR (Γ_n) = 1.5	0.505 (0.102, 94.2%)	0.446 (0.099, 91.8%)	0.394 (0.097, 81.8%)	0.349 (0.095, 61.6%)
OR (Γ_n) = 2.0	0.505 (0.102, 94.2%)	0.446 (0.099, 92.0%)	0.392 (0.097, 81.2%)	0.348 (0.094, 60.8%)
OR (Γ_n) = 2.5	0.505 (0.102, 94.2%)	0.446 (0.099, 91.6%)	0.392 (0.096, 81.2%)	0.348 (0.094, 59.8%)

Table 2.b. Average CACE estimates when assuming latent ignorability for missingness of outcomes in the presence of a covariate (SE and coverage in parentheses).

Deviation from CE	Deviation from SI (difference between π_{R_n} and π_{R_c})			
	0%	10%	20%	30%
OR (Γ_n) = 1.0	0.506 (0.106, 92.6%)	0.507 (0.100, 95.2%)	0.505 (0.095, 94.0%)	0.506 (0.090, 94.6%)
OR (Γ_n) = 1.5	0.421 (0.111, 88.8%)	0.424 (0.104, 89.4%)	0.427 (0.098, 89.2%)	0.432 (0.093, 89.2%)
OR (Γ_n) = 2.0	0.365 (0.115, 77.6%)	0.370 (0.108, 78.8%)	0.375 (0.101, 76.2%)	0.383 (0.095, 76.0%)
OR (Γ_n) = 2.5	0.324 (0.119, 68.0%)	0.330 (0.111, 67.8%)	0.337 (0.104, 66.4%)	0.347 (0.097, 62.6%)

5 CONCLUSION

It was demonstrated in this study that causal effect estimates can be quite sensitive to violation of the exclusion restriction in outcome missingness, which is less known than the impact of violating the exclusion restriction in observed outcomes (Angrist et al., 1996; Jo, 2002). The results call for reconsideration of the notion that latent ignorability is always advantageous or at least harmless, and for more careful decisions in applying ignorability or latent ignorability models.

The relative benefit of models assuming ignorability and latent ignorability depends on the degree of deviation from ignorability and compound exclusion. As shown in the simulation study, in some situations, the impact of deviation from ignorability may outweigh the impact of deviation from compound exclusion, resulting in more biased causal effect estimates in models assuming standard ignorability than in models assuming latent ignorability. In other situations, the impact of deviation from compound exclusion may outweigh the impact of deviation from ignorability, resulting in more biased causal effect estimates in models assuming latent ignorability than in models assuming standard ignorability.

Evaluating scientific plausibility of compound exclusion and latent ignorability is often difficult due to the various known and unknown possibilities. To empirically examine the plausibility of standard ignorability and compound exclusion, it is useful to conduct sensitivity analyses of models imposing different combinations of these assumptions, and ultimately to relax both assumptions. These investigations can be carried out by relaxing compound exclusion (e.g., Frangakis et al., 2002; Hirano et al., 2000), or by employing alternative structural assumptions (e.g., Jo, in press).

However, the applicability of these alternative models has been explored only in the context of the standard exclusion restriction. More research is necessary to examine the efficiency of these approaches in exploring potential violation of compound exclusion under latent ignorability.

In examining the bias mechanism in CACE estimation, this study assumed a 50% compliance rate and a 50% average response rate in the control conditionm. Therefore, sensitivity of CACE estimates and the relative benefit of assuming standard ignorability and latent ignorability may vary in settings different from those employed in the current study. More in-depth examination is necessary considering a broad range of parameter values (including compliance rates and response rates) to systematically investigate sensitivity of causal effects under different missing-data mechanisms. This study also assumed in its investigation that the exclusion restriction holds in observed outcomes. However, the exclusion restriction can be violated for both observed outcomes and missingness of the outcomes. Therefore, to gauge the overall impact of violating compound exclusion, one needs to consider deviations from the exclusion restriction in both outcomes and missingness of outcomes. It would be interesting to study how these violations interplay and simultaneously affect causal effect estimates. Pre-treatment covariates were also considered as sources of information that may decrease sensitivity of causal effect estimates to model misspecifications. This study examined only the case, in which covarites are good predictors of compliance. However, in principle, covariates may also be associated with outcomes and missingness of outcomes. It is not well known how this information affects sensitivity of causal effects under different missing-data mechanisms. Further study

is needed to explore various factors associated with insensitivity of ignorability and latent ignorability models to model misspecifications.

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444-455.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*, 225-246.

Dempster, A., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

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

Frangakis, C. E. & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics, 58, 21-29.*

Frangakis, C. E., Rubin, D. B., & Zhou, X. H. (2002). Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms. *Biostatistics, 3, 147-164.*

Hirano, K., Imbens, G. W., Rubin, D. B. & Zhou, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics, 1*, 69-88.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945-970

Imbens, G. W. & Angrist. J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*, 467-476.

Imbens, G. W. & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with non-compliance. *The Annals of Statistics, 25*, 305-327.

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. *Statistics in Medicine, 21*, 3161-3181.

Jo, B. (in press). Estimating intervention effects with noncompliance: Alternative model specifications. Forthcoming in the *Journal of Educational and Behavioral Statistics* (with discussion and rejoinder).

Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley & Sons

Little, R. J. A. & 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*, 147-159.

McLachlan, G. J. & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: Wiley Series.

Muthén, B. O. & Brown, C. H. (2001). Non-ignorable missing data in a general latent variable modeling framework. Draft.

Muthén, L. K. & Muthén, B. O. (1998-2001). *Mplus user's guide.* Los Angeles: Muthén & Muthén.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics, 6*, 34-58.

Rubin, D. B. (1980). Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association, 75*, 591-593.

Rubin, D. B. (1990). Comment on "Neyman (1923) and causal inference in experiments and observational studies." *Statistical Science, 5*, 472-480.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Tanner, M. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions.* New York: Springer.

West, S. G. & Sagarin, B. J. (2000). Participant selection and loss in randomized experiments. In L. Bickman (Ed.), *Research Design* (pp. 117-154). Thousand Oaks, CA: Sage.