

Methods for clustered encouragement design studies with noncompliance and missing data

LESLIE TAYLOR*

*Biostatistics Unit, Health Services Research and Development Center of Excellence,
VA Puget Sound Health Care System, 1100 Olive Way #1400,
Seattle, WA 98101, USA
taylorl@u.washington.edu*

XIAO-HUA ZHOU

*Department of Biostatistics, University of Washington, Seattle, WA 98195, USA and
Biostatistics Unit, Health Services Research and Development Center of Excellence,
VA Puget Sound Health Care System, 1100 Olive Way #1400, Seattle, WA 98101, USA*

SUMMARY

Encouragement design studies are particularly useful for estimating the effect of an intervention that cannot itself be randomly administered to some and not to others. They require a randomly selected group receive extra encouragement to undertake the treatment of interest, where the encouragement typically takes the form of additional information or incentives. We consider a “clustered encouragement design” (CED), where the randomization is at the level of the clusters (e.g. physicians), but the compliance with assignment is at the level of the units (e.g. patients) within clusters. Noncompliance and missing data are particular problems in encouragement design studies, where encouragement to take the treatment, rather than the treatment itself, is randomized. The motivating study looks at whether computer-based care suggestions can improve patient outcomes in veterans with chronic heart failure. Since physician adherence has been inadequate, the original study focused on methods to improve physician adherence, although an equally important question is whether physician adherence improves patient outcomes. Here, we re-analyze the data to determine the effect of physician adherence on patient outcomes. We propose causal inference methodology for the effect of a treatment versus a control in a randomized CED study with all-or-none compliance at the unit level. These methods extend the current approaches to account for non-ignorable missing data and use an alternative approach to inference using multiple imputation methods, which have been successfully applied to a wide variety of missing data problems and have recently been applied to the potential outcomes framework of causal inference (Taylor and Zhou, 2009b).

Keywords: Causal inference; Clustered encouragement design; Complier average causal effect; Missing data; Multiple imputation; Noncompliance; Nonresponse; Principal stratification.

*To whom correspondence should be addressed.

1. INTRODUCTION

For treatments that cannot be randomized, for ethical reasons or for lack of feasibility, it may be possible to randomize encouragement to the treatment of interest, which would increase use of treatment in the intervention arm while leaving the control arm treatment use unchanged. If the effect of encouragement is of interest, intention-to-treat (ITT) analyses are appropriate. But if interest lies in the effect of the treatment itself on outcome, then methods that focus on compliant subjects are more appropriate.

Our methods utilize a causal inference approach, referred to as principal stratification (Frangakis and Rubin, 2002), which focuses on the potential outcomes among the subpopulation of compliers, who are not fully identifiable from the observed data. Alternative methods of modeling noncompliance, such as conditioning on the potential outcomes in selection models (Heitjan, 1999) or the use of structural equation models (Robins and others, 1999), are not considered here. For research on nonignorable missing data methods and treatment noncompliance under the principal stratification framework, Frangakis and Rubin (1999) developed methods for encouragement designs using method of moments for inference; Taylor and Zhou (2009b) developed methods for this setting using a multiple imputation framework for inference. Barnard and others (1988) described methods that extend this template to allow for missing covariates and multivariate outcomes using a basic multiple imputations approach to inference, although they stopped short of implementing their methodology. Barnard and others (2003) later extended this framework to a more complicated form of crossover noncompliance (i.e. control patients could receive the intervention and intervention patients could receive the control) and used a Bayesian framework for inference. For research on encouragement design studies, Hirano and others (2000) developed methods for a binary outcome in a setting where outcomes are completely observed, relaxing the exclusion restrictions (often used to develop and identify estimands of interest) using a Bayesian framework for inference. Ten Have and others (2004) extended their approach to linear models for the continuous outcome. For research on CED studies, Frangakis and others (2002) developed methods in a setting where outcomes are completely observed, using a Bayesian framework for inference.

We propose to extend the methods of Frangakis and others (2002) for CED studies by incorporating nonignorable outcome nonresponse and by utilizing a multiple imputation approach to inference, which has been successfully applied to a wide variety of missing data problems, but has only recently been applied to the potential outcomes framework of causal inference. Taylor and Zhou (2009b) developed multiple imputation methods for the setting of crossover noncompliance and outcome nonresponse for a simple binary outcome without covariates or clustering and compared these methods to existing likelihood and moment methods. We extend these multiple imputation methods on binary outcome data to include clustering by physician and patient covariates and apply these methods to the guideline treatment study.

We begin with a description of the study that motivates our methods in Section 2. Section 3 defines the causal inference notation, assumptions, and parameters of interest. Section 4 defines the proposed models, and Section 5 describes the multiple imputation framework for inference. In Section 6, these methods are applied to data from a guideline treatment study previously studied by Subramanian and others (2004).

2. MOTIVATIONAL STUDY

These methods are motivated by a study in veterans with chronic heart failure (Subramanian and others, 2004). Chronic heart failure is a highly prevalent condition associated with serious morbidity, intense levels of health use, and shortened survival, where approximately two-thirds of patients die within 5 years of initial hospitalization (Ashton and others, 2000). Veterans Affairs (VA) treatment guidelines in this population have been created to improve quality of health care, lessen clinical practice variation, and reduce costs (Audet and others, 1990; Institute of Medicine, 1992). These guidelines consist of an algorithm that allows a clinician to follow a linear approach to evaluate and develop management strategies for the

patient. These guideline rules are typically programmed into a locally developed decision support system and the workstations used by physicians.

The original study focused on methods to improve physician adherence through use of computer-based care suggestions, which involved sending the physician patient-specific electronic guideline treatment suggestions before each patient visit. Physicians were randomly assigned to receive either care suggestions generated with electronic medical data as well as symptom data obtained from questionnaires mailed to patients within 2 weeks of scheduled outpatient visits (intervention physicians) or care suggestions generated with electronic medical data alone (control physicians, where the control suggestions were considered a type of “standard of care” in these facilities). Since physician adherence had been inadequate, the original study focused on methods to improve physician adherence, although an equally important question is whether increased physician adherence improves patient outcomes. Under this study aim, which is the focus of this paper, physician-randomized intervention (e.g. physician encouragement to administer the guideline treatment) is a surrogate to induce the target treatment of interest (e.g. physician adherence), and thus there is a large amount of noncompliance.

3. POTENTIAL OUTCOMES FRAMEWORK

Causal inference, a growing area of statistical research, can be viewed as the extraction of information on the comparison between therapies or policies. An important and much-recognized statistical framework for causal inference is the one based on potential outcomes, which dates back to Neyman and Fisher (Neyman, 1923; Rubin, 1990), and is often referred to as the Rubin Causal Model (Holland, 1986). The approach to adjusting for noncompliance within this framework is a special case of the principal stratification framework defined by Frangakis and Rubin (2002).

The role of time is important in this framework since variables are divided into two categories: pretreatment variables whose values are determined prior to treatment exposure and posttreatment variables whose values are determined after treatment exposure (Holland, 1986). All posttreatment variables are potentially affected by the treatment or control. So for subject i , each posttreatment outcome variable Y_i has two “potential outcomes”: one under treatment, denoted $Y_i(1)$, and one under control, denoted $Y_i(0)$. We interpret $Y_i(1)$ as the value of the response that would have been observed if subject i were assigned the treatment, and $Y_i(0)$ as the value of the response that would have been observed on the same subject i if assigned control. It is impossible to observe both $Y_i(1)$ and $Y_i(0)$ since the subject is randomized to either treatment or control, but a statistical solution is found in working with average differences in potential outcomes across populations. In our research setting, posttreatment variables, for which potential outcomes will be specified, include physician adherence, patient outcomes, and response indicators.

3.1 Setting and notation

The setting consists of $i = 1, \dots, N$ patients, where patient i has physician Q_i for $Q_i = 1, \dots, M \leq N$. Then, $Z_i = 1$ if patient i 's physician is randomized to encouragement with medical records plus symptoms data, and 0 if patient i 's physician is randomized to encouragement with medical records data alone. Since randomization is by physician cluster, $Z_i = Z_j$ for $Q_i = Q_j$. Under the stable unit treatment value assumption (SUTVA), the potential physician adherence indicator $D_i(Z_i) = 1$ if patient i 's physician Q_i adheres to guideline treatment under Z_i , and $D_i(Z_i) = 0$ otherwise.

Since patients had multiple visits over the 12 months of the study, we collapsed physician adherence to a binary variable, where a patient i 's physician Q_i is considered to have adhered if they provided guideline treatment at least once during the 12 months. We define guideline treatment to be treatment

Table 1. *Notation*

Notation	Specifics	General description
Z_i	1 if patient i assigned treatment 0 if patient i assigned control	Treatment assignment indicator
X_i	Vector of indicators of baseline NYHA status	Patient covariates
Q_i	q for patient i 's physician	
$D_i(Z_i)$	1 if patient i 's physician adhered under assignment Z_i 0 if patient i 's physician did not adhere under assignment Z_i	Potential outcome formulation of treatment receipt
D_i		Treatment receipt indicator under observed assignment
C_i	n if $D_i(0) = 0$ and $D_i(1) = 0$ c if $D_i(0) = 0$ and $D_i(1) = 1$ a if $D_i(0) = 1$ and $D_i(1) = 1$ d if $D_i(0) = 1$ and $D_i(1) = 0$	Compliance type principal stratum: n = never-taker, c = complier, a = always-taker; d = defier
$Y_i(Z_i)$	Binary outcome of interest under assignment Z_i	Potential outcome formulation of the outcome of interest
Y_i		Binary outcome of interest under observed assignment
$R_i(Z_i)$	1 if $Y_i(Z_i)$ would be observed 0 if $Y_i(Z_i)$ would not be observed	Potential outcome formulation of the response indicator for Y_i
R_i		Response indicator for Y_i under observed assignment

based on medical records and patient-specific symptom data. We have this data since both intervention and control suggestions were generated for all patients at all visits in the study; all future references to physician adherence will refer to this definition.

Under SUTVA, we define $Y_i(Z_i)$ to be the binary potential outcome of interest, when patient i 's physician is assigned Z_i . In the guideline treatment study, Y_i is a binary indicator of whether a patient's physical status improved over the 12 months of the study as measured by their New York Heart Association (NYHA) score. Let $R_i(Z_i)$ be the potential indicator for response, equal to 1 if $Y_i(Z_i)$ would be observed, and 0 if not, when patient i 's physician is assigned Z_i . Let X_i be a vector of fully observed covariates for patient i . In the application, we adjust for baseline NYHA class since there were unequal numbers at baseline between treatment groups, where X_i corresponds to binary indicators of baseline NYHA classes. There were no recorded physician-level covariates in the guideline treatment study. Table 1 provides a summary of all notation.

3.2 Assumptions

Our methods assume the SUTVA which states that a patient's treatment assignment does not affect other patients' outcomes. If we consider a physician with two patients, j and k , then SUTVA implies that how the physician treats patient j should not influence the outcome of patient k . In the application, the electronic reminders were sent prior to each patient visit and were patient specific (i.e. based on a specific patient's medical records, as well as, for intervention physicians, the patient's symptom information). So it is reasonable to expect the physician's responses to be specific to each patient, and independent of the physician's treatment of his or her other patients, which satisfies SUTVA. Although physician responses may be correlated within physician, which our methods take into account. The SUTVA then limits the number of potential outcomes by allowing us to write potential outcomes as functions of Z_i rather than of the vector $\mathbf{Z}$. We can then define the ITT causal effect of Z on Y as $E[Y_i(1) - Y_i(0)]$.

In the guideline treatment study, if the patient's physician does not adhere when encouraged with medical data and patient-specific symptom data (i.e. the intervention), then we assume the patient's physician would not adhere when encouraged with medical data alone (i.e. the control). This is the "monotonicity assumption" (Imbens and Angrist, 1994), which rules out the existence of defiers, making compliance type observable when $Z_i \neq D_i$.

Unintentional missing data create problems since we are typically interested in making inference on the entire target population, not just the portion of the population that would provide responses. We choose a less restrictive assumption where we assume the missing data are nonignorable, which means that non-response is related in some partially unknown way to the missing data, even after adjusting for covariates known for responders and nonresponders. Specifically in our setting, missingness depends on latent compliance type (defined in Section 3.3), which is missing for patients where $Z_i = D_i$ (under monotonicity); this assumption is called "latent ignorability" (Frangakis and Rubin, 1999) and formally, it assumes that within each latent principal compliance type as well as covariates and physician cluster, potential outcomes and associated potential response indicators are independent, or $\Pr[R_i(0)|Y_i(0), C_i, X_i, Q_i] = \Pr[R_i(0)|C_i, X_i, Q_i]$ and $\Pr[R_i(1)|Y_i(1), C_i, X_i, Q_i] = \Pr[R_i(1)|C_i, X_i, Q_i]$. We condition on covariates and physician clusters for two reasons: first, our methods and inference require conditioning on physician cluster since this is the level of the randomization and second, the assumption of ignorability is more plausible after conditioning on important covariates.

We do not assume any exclusion restrictions which hold that for certain subpopulations, treatment assignment does not affect outcomes or response behaviors. These restrictions would be violated if, for example, the electronic reminder prompts the physician to take measures beyond the suggested guidelines, like suggesting that the patient avoid unnecessary exposure to certain things. Without the exclusion restrictions, the parameters in Section 3.4 are "weakly identifiable" in the sense that there is no unique maximum likelihood (ML) estimator, although the posterior distribution is proper (Hirano and others, 2000). Identification under Bayesian or multiple imputation methods is then based entirely on a carefully thought-out prior distribution.

3.3 Compliance principal strata

We then use the method of principal stratification (Frangakis and Rubin, 2002), which divides the population into four compliance principal strata, as determined by the value of the vector of potential physician adherence indicators $[D_i(0), D_i(1)]$, where $C_i = n$ (never-taker) if $D_i(0) = D_i(1) = 0$, $C_i = c$ (complier) if $D_i(0) = 0$ and $D_i(1) = 1$, $C_i = a$ (always-taker) if $D_i(0) = D_i(1) = 1$, and $C_i = d$ (defier) if $D_i(0) = 1$ and $D_i(1) = 0$. Compliance type is not identified from the observable data on physician adherence D without additional assumptions. In particular, since patients are randomized to either the intervention or the control, we only observe $D_i(1)$ for patients assigned to the intervention, and $D_i(0)$ for patients assigned to the control. But there is some information on compliance type from the observed treatment assignment and physician adherence. For example, patients with physicians assigned the control arm where the physician was not guideline adherent (i.e. $Z_i = D_i = 0$) are a mixture of never-takers and compliers. Similarly, patients with physicians assigned to the intervention arm where the physician was adherent (i.e. $Z_i = D_i = 1$) are a mixture of always-takers and compliers. Note that conditioning on fully observed pretreatment covariates could substantially limit the extent of the relationship between $D_i(1)$ and $D_i(0)$ that is unobserved, depending on the strength of the covariates in predicting $D_i(1)$ and $D_i(0)$ (Zhang and Rubin, 2003; Rubin, 2004). Principal strata have some important properties. First, they are not affected by assigned treatment, so we can adjust for them as we do for baseline covariates (Frangakis and Rubin, 2002). Second, any comparisons made between potential outcomes within the principal strata are well-defined causal effects (Frangakis and Rubin, 2002), allowing us to define estimands for the causal treatment effect that appropriately account for noncompliance.

3.4 Defining the causal estimands of interest

We define the ITT effect as $ITT = \sum_{t \in \{n, a, c, d\}} E[Y_i(1) - Y_i(0) | C_i = t] \Pr(C_i = t) = \sum_{t \in \{n, a, c, d\}} ACE_t \Pr(C_i = t)$, where ACE_t is the average causal effect of randomized treatment Z on outcome Y for the subpopulation of compliance type t . Under monotonicity, defiers do not exist, and the never-takers and always-takers, by definition of these groups, do not carry information about the comparison between treatments with respect to finding the causal effect of treatment on outcome (Sommer and Zeger, 1991; Sheiner and Rubin, 1995). Therefore, we focus on the subpopulation of compliers and define the complier average causal effect (CACE) to be $CACE = E[Y_i(1) - Y_i(0) | C_i = c]$, which is the average treatment effect among the subpopulation of compliers. In the application, the CACE is the comparison of randomized groups on patient NYHA improvement among patients whose physician would adhere when assigned the intervention and would not adhere when assigned the control. Put simply, the CACE is the as-treated effect of physician adherence on patient NYHA improvement.

4. PARAMETRIC MODELS

Since randomization is at the level of the physician, randomized assignment is ignorable only after conditioning on physician cluster. Therefore, any inference on potential outcomes must condition on physician cluster which may lessen the precision since no physician has patients in both arms (Frangakis and others, 2002). Under latent ignorability, we model the joint distribution of $Y_i(Z_i)$, $R_i(Z_i)$, and C_i , conditionally on X and Q , as $\Pr[Y_i(Z_i) | C_i, Z_i, X_i, Q_i] \Pr[R_i(Z_i) | C_i, Z_i, X_i, Q_i] \Pr[C_i | Z_i, X_i, Q_i]$.

We model the compliance principal strata as a multinomial logit model, or $\Pr(C_i = t | X_i = x, Q_i = q, \theta) = \Psi(t, x, b_q^{(C)}, \psi)$, where, for $x \equiv (x_1, x_2, x_3)$, $\psi_t \equiv (\psi_{t1}, \psi_{t2}, \psi_{t3})$, $b_q^{(C)} \equiv (b_{qn}^{(C)}, b_{qc}^{(C)}, b_{qa}^{(C)})$, and $t \in \{n, c, a\}$,

$$\Psi(t, x, b_q^{(C)}, \psi) = \frac{\exp(\psi_{t0} + \psi_t x + b_{qt}^{(C)})}{\sum_{v \in \{n, c, a\}} \exp(\psi_{v0} + \psi_v x + b_{qv}^{(C)})}.$$

We set $\{\psi_{n0}, \psi_{n1}, \psi_{n2}, \psi_{n3}\} \equiv (0, 0, 0, 0)$, as well as $b_{qn}^{(C)} \equiv 0$ for $q = 1, \dots, M$, to normalize the probabilities. Here, $\{\psi_{c0}, \psi_{c1}, \psi_{c2}, \psi_{c3}\}$ and $\{\psi_{a0}, \psi_{a1}, \psi_{a2}, \psi_{a3}\}$ model the association between compliance principal strata and covariates, and $b_q^{(C)}$ model the association between compliance principal strata and physician cluster.

We model the binary potential outcomes as logistic regressions or $\Pr[Y_i(z) = 1 | C_i = t, X_i = x, Z_i = z, Q_i = q, \theta] = \Upsilon(x, \eta_{tz}, b_q^{(Y)})$ where, for $\eta_{tz} \equiv \{\eta_{tz1}, \eta_{tz2}, \eta_{tz3}\}$,

$$\Upsilon(x, \eta_{tz}, b_q^{(Y)}) = \frac{\exp\{\eta_{tz0} + \eta_{tz} x + b_q^{(Y)}\}}{1 + \exp\{\eta_{tz0} + \eta_{tz} x + b_q^{(Y)}\}}.$$

We let $\eta_{t01} = \eta_{t11} \equiv \eta_{0.1}$, $\eta_{t02} = \eta_{t12} \equiv \eta_{0.2}$, and $\eta_{t03} = \eta_{t13} \equiv \eta_{0.3}$ for $t \in \{n, c, a\}$, which reduces the number of parameters. Although it does not complicate the methodology to allow separate coefficients, we did not have a strong a priori belief that the baseline covariate's effect on potential outcomes would vary by compliance type or treatment assignment. Given the sample size in this study, this should help increase precision in the estimates of interest.

We model the binary potential outcome responses as logistic regressions, or $\Pr[R_i(z) = 1 | C_i = t, X_i = x, Z_i = z, Q_i = q, \theta] = \Xi(x, \gamma_{tz}, b_q^{(R)})$ where, for $\gamma_{tz} \equiv \{\gamma_{tz1}, \gamma_{tz2}, \gamma_{tz3}\}$,

$$\Xi(x, \gamma_{tz}, b_q^{(R)}) = \frac{\exp\{\gamma_{tz0} + \gamma_{tz} x + b_q^{(R)}\}}{1 + \exp\{\gamma_{tz0} + \gamma_{tz} x + b_q^{(R)}\}}.$$

Similar to the outcome model parameters, we let $\gamma_{t01} = \gamma_{t11} \equiv \gamma_{0,1}$, $\gamma_{t02} = \gamma_{t12} \equiv \gamma_{0,2}$, and $\gamma_{t03} = \gamma_{t13} \equiv \gamma_{0,3}$ for $t \in \{n, c, a\}$.

4.1 Likelihood

We define the vector of all parameters to be θ . Let $\zeta(t)$ be the subset of patients with complier type t , and let $\delta(z, d)$ be the subset of patients with observed ($Z_i = z, D_i = d$). The observed likelihood takes a complicated form and is defined in Section 1 of the supplementary material available at *Biostatistics* online. However, the likelihood based on the complete data $\{\mathbf{C}, \mathbf{Z}, \mathbf{D}, \mathbf{Y}, \mathbf{R}, \mathbf{X}, \mathbf{Q}\}$ takes a much simpler form, where

$$\begin{aligned} L_{\text{comp}}(\theta | \mathbf{C}, \mathbf{Z}, \mathbf{D}, \mathbf{Y}, \mathbf{R}, \mathbf{X}, \mathbf{Q}) = & \prod_{i \in \zeta(c)} \left\{ \Upsilon(x, \eta_{cZ_i}, b_q^{(Y)})^{Y_i} (1 - \Upsilon(x, \eta_{cZ_i}, b_q^{(Y)}))^{(1-Y_i)} \right. \\ & \times \Xi(x, \gamma_{cZ_i}, b_q^{(R)})^{R_i} (1 - \Xi(x, \gamma_{cZ_i}, b_q^{(R)}))^{(1-R_i)} \\ & \left. \times \Psi(c, x, b_q^{(C)}, \psi) \right\} \\ & \prod_{i \in \zeta(n)} \left\{ \Upsilon(x, \eta_{nZ_i}, b_q^{(Y)})^{Y_i} (1 - \Upsilon(x, \eta_{nZ_i}, b_q^{(Y)}))^{(1-Y_i)} \right. \\ & \times \Xi(x, \gamma_{nZ_i}, b_q^{(R)})^{R_i} (1 - \Xi(x, \gamma_{nZ_i}, b_q^{(R)}))^{(1-R_i)} \\ & \left. \times \Psi(n, x, b_q^{(C)}, \psi) \right\} \\ & \prod_{i \in \zeta(a)} \left\{ \Upsilon(x, \eta_{aZ_i}, b_q^{(Y)})^{Y_i} (1 - \Upsilon(x, \eta_{aZ_i}, b_q^{(Y)}))^{(1-Y_i)} \right. \\ & \times \Xi(x, \gamma_{aZ_i}, b_q^{(R)})^{R_i} (1 - \Xi(x, \gamma_{aZ_i}, b_q^{(R)}))^{(1-R_i)} \\ & \left. \times \Psi(a, x, b_q^{(C)}, \psi) \right\}. \end{aligned}$$

Since the complete data likelihood is easier to work with than the observed data likelihood, we can utilize missing data methods such as the Gibbs sampler (Geman and Geman, 1984).

4.2 Prior selection

We perform two separate analyses using two types of priors that are computationally tractable for this application. For the first type, we use a prior distribution that is proper but diffuse in order to be relatively uninformative for our application. Given the physician-specific parameters, we posit the following independent prior distributions: $\eta \sim N(0, \mathbf{I}_\xi)$, $\gamma \sim N(0, \mathbf{I}_\xi)$, $\psi \sim N(0, \mathbf{I}_\xi)$, and $(b_q^{(Y)}, b_q^{(R)}, b_q^{(C)}, b_{qc}^{(C)}, b_{qa}^{(C)}) | V_b \sim N(0, V_b)$ independently across physicians $q = 1, \dots, M$. $\mathbf{I}$ is the identity matrix and ξ is the inflation factor that we set. We assume that the inverse of V_b has a Wishart distribution with scale matrix $(\alpha R)^{-1}$ with $\alpha = 10$ degrees of freedom.

For the second prior, since the observed likelihood involves a mixture over a large amount of missing data, we use a prior with conjugate form. This more informative prior can be fit using the same model but adding artificial patients to the data set (Hirano and others, 2000). Prior information is expressed by attaching weights to the artificial patients' contribution to the likelihood, where weights are based on the combinations of cluster indicators, covariates, compliance groups, randomized assignment, outcomes, and

observed response indicators. We use a prior that corresponds to adding 30 observations to the likelihood. This corresponds to 10 additional observation for each compliance type, where the 10 observations are then split into $10/(32Q)$ observations for each of the $32Q$ combinations of $(Z_i, Y_i, R_i, X_{i1}, X_{i2}, X_{i3}, Q)$. Formally, the prior distribution is as follows:

$$\begin{aligned}
 p(\theta) \propto & \prod_{q=1}^Q \times \prod_{t \in \{n, c, a\}} \times \prod_x \prod_{z=0,1} \prod_{y=0,1} \prod_{r=0,1} \left\{ \Upsilon(x, \eta_{cZ_i.}, b_q^{(Y)})^{Y_i} (1 - \Upsilon(x, \eta_{cZ_i.}, b_q^{(Y)}))^{(1-Y_i)} \right. \\
 & \times \Xi(x, \gamma_{a1.}, b_q^{(R)})^{R_i} (1 - \Xi(x, \gamma_{a1.}, b_q^{(R)}))^{(1-R_i)} \\
 & \left. \times \Psi(a, x_1, x_2, b_{qc}^{(C)}, b_{qa}^{(C)}, \psi) \right\}^{10/(32Q)} \\
 & \times \text{MVN}(0, V_b) \times \text{InvWis}(R, \alpha),
 \end{aligned}$$

where $\text{MVN}(0, V_b)$ is the multivariate normal distribution for b_q with mean vector $\mathbf{0}$ and variance V_b , and InvWis is the inverse Wishart distribution we assume for the distribution of V_b , with hyperparameters $(\alpha R)^{-1}$ and α for the scale matrix and degrees of freedom, where, for the scale matrix, we use an approximation of V_b based on the posterior distribution under the uninformative prior distribution described above, with $\alpha = 10$.

5. FRAMEWORK FOR MULTIPLE IMPUTATION

In the principal stratification framework, existing methods for estimation of the treatment effect include Bayesian (Imbens and Rubin, 1997; Hirano and others, 2000; Frangakis and others, 2002), likelihood (O'Malley and Normand, 2005; Zhou and Li, 2005), and moment methods (Frangakis and Rubin, 1999; Levy and others, 2004; O'Malley and Normand, 2005; Taylor and Zhou, 2009a). An interesting alternative is multiple imputation (Rubin, 1987), a method that has been successfully applied to a wide variety of missing data problems but has only recently been applied to the potential outcomes setting of causal inference (Taylor and Zhou, 2009b).

Taylor and Zhou (2009b) provide a detailed list of the advantages of multiple imputation (MI) in this causal inference setting, which include the fact that neither the monotonicity nor the compound exclusion restrictions, which have been used to develop and justify estimation procedures under ML and moment methods, are essential for the MI analysis, and violations of these assumptions can be addressed. In addition, MI can be used in virtually any setting, whereas ML is much more restricted in its applications, requiring specially designed expectation-maximization algorithms and difficult analytic or numerical integration on the log likelihood for intervals.

If we partition the complete data M into the observed data, $M_{\text{obs}} = (Y_{\text{obs}}, Z, D, R, X, Q)$, and the missing data, $M_{\text{mis}} = (Y_{\text{mis}}, C_{\text{mis}})$, we then draw from the joint distribution $P(M_{\text{mis}}, \theta | M_{\text{obs}}) \equiv P(Y_{\text{mis}}, C_{\text{mis}}, \theta | Y_{\text{obs}}, Z, D, R, X, Q)$ by recursively iterating between the draws from $P(Y_{\text{mis}} | C_{\text{mis}}, Y_{\text{obs}}, Z, D, R, X, Q, \theta)$, $P(C_{\text{mis}} | Y_{\text{obs}}, Z, D, R, X, Q, \theta)$, and $P(\theta | Y_{\text{obs}}, Z, D, R, X, Q)$. We use a Gibbs sampler (Geman and Geman, 1984) and a Metropolis–Hastings algorithm (Metropolis and others, 1953; Hastings, 1970) for computation of all reported results, where the details are described in Section 2 of the supplementary material available at *Biostatistics* online.

For our setting, if the percent of missing information was even as high as 90%, then an estimate based on $m = 10$ imputations would tend to have a standard error only 1.04 times as large as the estimate with $m = \infty$ imputations; therefore 10 imputations is adequate for our setting. The next step, or the analysis stage, involves calculating the complete data estimates of the estimands of interest using complete data methods.

Since we are interested in the marginal mean, but would like to use likelihood-based methods at the analysis stage (which were also used in the imputation stage), we chose a marginalized random effect model (Heagerty and Zeger, 2000), which uses a conditional approach but averages over random effects. This approach embraces the interpretation and robustness of regression coefficients from the marginal model while retaining the likelihood inference capabilities and flexible dependence specifications from a random effects model (Griswold and Zeger, 2004). This model can be implemented in existing software (like SAS) by formulating the marginal models through conditional specifications, which facilitate estimation with computation solutions already in place (Griswold and Zeger, 2004). See Section 3 of the supplementary materials section available at *Biostatistics* online for a more detailed description of the marginalized random effects models used at the analysis stage.

For each complete data set, we obtain parameter estimates for the CACE and compliance type, $(\hat{Q}^1, \hat{Q}^2, \dots, \hat{Q}^m)$, along with their approximate variances, $(U^1, U^2, \dots, U^m)$, based on the second derivative matrix of the likelihood function. Similar methods are used to obtain estimates for all estimands of interest. We then combine results from the completed data sets using the rules defined in Rubin (1987).

6. APPLICATION TO THE GUIDELINE TREATMENT STUDY

The study was done among primary care physicians caring for outpatients with heart failure at two VA medical centers. Although subjects were not eligible to participate in the original study if their physicians did not expect them to survive 1 year (Subramanian and others, 2004), death data were not included in this analysis. Since outcomes for patients who die are, by definition, nonexistent, it is a limitation of this analysis that outcomes may have been imputed for patients who died during the 12 months of the study. A discussion of how to adjust for censoring due to death using the principal stratification framework can be found in Zhang and Rubin (2003) and Rubin (2006).

Table 2 provides patient characteristics and baseline information, where 38% of patients under the control ($Z_i = 0$) had NYHA improvement over the 12 months of the study, and 45% of patients whose physician was assigned the intervention had improvement. A substantial number of veterans with heart failure were not receiving care in accordance with accepted local guidelines. Overall, less than half the

Table 2. Patient characteristics for the guideline treatment study

	Intervention assignment	Control assignment	Overall
Physicians, N	38	36	74
Patients, N	322	307	629
Received treatment ($D_i = 1$), N (%)	157 (47)	114 (36)	271 (42)
Baseline NYHA Class, N (%)			
I—No symptoms or limitations	36 (11)	61 (20)	97 (16)
II—Mild symptoms and limitations	73 (23)	64 (21)	137 (22)
III—Marked limitations due to symptoms	161 (50)	147 (48)	308 (49)
IV—Severe symptoms and limitations, bedbound	52 (16)	35 (11)	87 (14)
NYHA improvement ($Y_i^{\text{obs}} = 1$), N (%)	103 (45)	89 (38)	192 (41)
Among those receiving treatment	57 (48)	35 (39)	92 (44)
Among those not receiving treatment	46 (41)	54 (38)	100 (40)
Observed NYHA status ($R_i = 1$), N (%)	231 (72)	232 (76)	463 (74)
Among those receiving treatment, N (%)	119 (78)	91 (80)	210 (79)
Among those not receiving treatment, N (%)	112 (66)	141 (73)	253 (70)

patients received guideline treatment at least once during the 12 months of this study. Approximately half the patients, whose physicians were assigned the intervention, did not receive guideline treatment; these patient–physician pairs must be never-takers (or, more descriptively, “never-guideline-treatment-receivers”); similarly there were 36% of patient–physician pairs in the control arm who must be always-takers (or “always-guideline-treatment-receivers”). The remaining patient–physician pairs in the control arm were either compliers or never-takers, and those remaining in the intervention arm were either compliers or always-takers.

For an initial examination of the data, we used the multiple imputation methods developed in [Taylor and Zhou \(2009b\)](#) on binary outcomes without covariates or clustering using $M = 10$ imputations. Results are reported in the first column of Table 3, where there was an estimated 12% (95% CI: $-67.6, 92.1$) NYHA improvement among the complier subpopulation. The results from adjusting for covariates are reported in the latter two columns of Table 3, corresponding to the two different priors used in the two analyses. We adjusted for baseline NYHA status (with three indicator variables for status II, III, and IV) since this was somewhat unbalanced at baseline, with patients in the intervention arm worse off (i.e. having higher NYHA scores at baseline) than control patients (Table 2).

Under the conjugate prior that adds 30 artificial patients to the likelihood, guideline treatment produced an estimated 4% (95% CI: $-25.2, 32.2$) improvement in NYHA status, where including covariates substantially lowered variability in the estimated treatment effect. Similar nonsignificant effects were found in the never-taker and always-taker subpopulations (.1% [95% CI: $-11.3, 11.5$]; 1.9% [95% CI: $-11.9, 15.6$], respectively). Similar results were found under the noninformative prior where the slightly negative treatment effect among the compliers had a wide confidence interval which included zero, making the negative result consistent with random fluctuation. The last six rows of Table 3 report results on response indicators for all models, where overall rates of outcome response were lowest among compliers and highest among always-takers.

Table 4 reports the estimated proportion within each compliance subpopulation by covariates. Overall, under the conjugate prior, 13% of the patient–physician pairs were compliers, 50% were never-takers

Table 3. Overview of NYHA improvement rates and response rates for compliance principal strata based on imputations from the posterior distribution

	Model without covariates	Model adjusting for covariates	
	Mean (SE) [95% CI]	Normal prior ($\xi = 1$) Mean (SE) [95% CI]	Conjugate prior Mean (SE) [95% CI]
$ACE_c \equiv CACE$	0.12 (0.36) [$-0.68, 0.92$]	-0.01 (0.15) [$-0.30, 0.28$]	0.04 (0.14) [$-0.25, 0.32$]
ACE_n	0.05 (0.08) [$-0.13, 0.22$]	0.01 (0.06) [$-0.10, 0.13$]	0.001 (0.06) [$-0.11, 0.12$]
ACE_a	0.07 (0.12) [$-0.18, 0.31$]	0.04 (0.09) [$-0.15, 0.22$]	0.02 (0.07) [$-0.12, 0.16$]
$E(Y_i(1) C_i = c)$	0.62 (0.36) [0.00, 1.00]	0.44 (0.15) [0.13, 0.75]	0.50 (0.09) [0.33, 0.68]
$E(Y_i(0) C_i = c)$	0.50 (0.35) [0.00, 1.00]	0.45 (0.14) [0.16, 0.74]	0.48 (0.09) [0.29, 0.65]
$E(Y_i(1) C_i = n)$	0.41 (0.05) [0.30, 0.52]	0.47 (0.04) [0.39, 0.56]	0.47 (0.06) [0.35, 0.58]
$E(Y_i(0) C_i = n)$	0.36 (0.06) [0.23, 0.50]	0.46 (0.05) [0.36, 0.56]	0.47 (0.05) [0.36, 0.57]
$E(Y_i(1) C_i = a)$	0.45 (0.10) [0.24, 0.67]	0.50 (0.05) [0.41, 0.59]	0.48 (0.06) [0.37, 0.60]
$E(Y_i(0) C_i = a)$	0.39 (0.05) [0.28, 0.49]	0.46 (0.07) [0.32, 0.60]	0.47 (0.07) [0.33, 0.60]
$E(R_i(1) C_i = c)$	0.60 (0.23) [0.08, 1.00]	0.58 (0.13) [0.31, 0.85]	0.65 (0.17) [0.29, 1.00]
$E(R_i(0) C_i = c)$	0.42 (0.29) [0.00, 1.00]	0.60 (0.16) [0.24, 0.95]	0.58 (0.24) [0.05, 1.00]
$E(R_i(1) C_i = n)$	0.63 (0.04) [0.55, 0.70]	0.65 (0.04) [0.57, 0.73]	0.65 (0.04) [0.57, 0.73]
$E(R_i(0) C_i = n)$	0.76 (0.07) [0.61, 0.90]	0.79 (0.07) [0.65, 0.94]	0.77 (0.08) [0.60, 0.94]
$E(R_i(1) C_i = a)$	0.83 (0.09) [0.62, 1.00]	0.82 (0.09) [0.61, 1.00]	0.80 (0.07) [0.65, 0.96]
$E(R_i(0) C_i = a)$	0.80 (0.04) [0.73, 0.87]	0.79 (0.04) [0.71, 0.87]	0.79 (0.04) [0.71, 0.87]

Table 4. Compliance principal strata rates by covariates

	Normal prior ($\zeta = 1$)	Conjugate prior (PS = 30)
	Estimate (SE) [95% CI]	Estimate (SE) [95% CI]
$\Pr(C_i = c)$	0.08 (0.06) [0.00, 0.22]	0.13 (0.05) [0.03, 0.25]
$\Pr(C_i = n)$	0.55 (0.08) [0.38, 0.71]	0.50 (0.08) [0.34, 0.66]
$\Pr(C_i = a)$	0.37 (0.08) [0.21, 0.52]	0.37 (0.09) [0.20, 0.54]
NYHA Class I: no limitations		
$\Pr(C_i = c)$	0.16 (0.17) [0.00, 0.53]	0.22 (0.14) [0.00, 0.50]
$\Pr(C_i = n)$	0.48 (0.12) [0.06, 0.89]	0.42 (0.19) [0.04, 0.80]
$\Pr(C_i = a)$	0.36 (0.09) [0.20, 0.60]	0.36 (0.09) [0.20, 0.54]
NYHA Class II: mild limitations		
$\Pr(C_i = c)$	0.05 (0.04) [0.00, 0.13]	0.15 (0.12) [0.00, 0.41]
$\Pr(C_i = n)$	0.62 (0.08) [0.47, 0.77]	0.52 (0.15) [0.22, 0.83]
$\Pr(C_i = a)$	0.33 (0.06) [0.21, 0.46]	0.33 (0.07) [0.17, 0.46]
NYHA Class III: marked limitations		
$\Pr(C_i = c)$	0.09 (0.07) [0.00, 0.22]	0.12 (0.09) [0.00, 0.32]
$\Pr(C_i = n)$	0.51 (0.08) [0.35, 0.68]	0.49 (0.11) [0.27, 0.71]
$\Pr(C_i = a)$	0.40 (0.05) [0.30, 0.51]	0.37 (0.05) [0.27, 0.46]
NYHA Class IV: severe limitations		
$\Pr(C_i = c)$	0.06 (0.07) [0.00, 0.20]	0.10 (0.09) [0.00, 0.29]
$\Pr(C_i = n)$	0.51 (0.08) [0.35, 0.68]	0.58 (0.12) [0.34, 0.83]
$\Pr(C_i = a)$	0.33 (0.08) [0.19, 0.48]	0.32 (0.08) [0.16, 0.48]

ζ = inflation factor set by analysis; PS = number of pseudo subjects in the conjugate prior.

(where a high proportion of never-takers was expected since guideline–treatment adherence is a known problem for physicians in this patient population), and 37% were always-takers. Similar results were found under the noninformative prior. Patients with higher NYHA baseline scores (where high scores are worse for the patient) tended to have slightly lower proportions of compliers and higher proportions of never-takers than patient with higher baseline NYHA scores.

In order to assess sensitivity of inference to the parameters in our priors, Table 5 displays results based on three different inflation factors ($\zeta = 5, 3$, and 0.5) under the normal priors, and three different numbers of pseudo subjects (PS) under the conjugate prior (PS = 9, 15, and 45). Inferences are similar both in the estimated treatment effect among compliance subpopulations and in the estimated proportions within each compliance subpopulation.

Since there was no significant improvement in NYHA status over the course of the study for the compliers, our results suggest that physician adherence to treatment guidelines in this setting does not necessarily improve patient outcomes. This implies a need for further research on management strategies for veterans with heart failure in order to provide them optimal care.

7. DISCUSSION

We developed methodology for cluster-randomized studies with individual-level noncompliance. The proposed methods improve upon current procedures for the clustered encouragement design by incorporating nonignorable missing data and by utilizing an alternative approach to inference using a multiple imputation framework which has potential advantages over existing methods of inference.

Future work could focus on continuous outcomes and incorporate additional physician adherence categories since, to make the observed data in the guideline treatment study conform to our template, we had

Table 5. Overview of NYHA improvement rates and compliance principal strata based on imputations from the posterior predictive distribution using different inflation factors (ξ) for the noninformative prior and different numbers of PS for the conjugate prior

Noninformative prior inflation factors, Mean (SE) [95% CI]	$\xi = 5$	$\xi = 3$	$\xi = 0.5$
ACE _c	0.01 (0.14) [−0.26, 0.28]	0.003 (0.20) [−0.38, 0.39]	0.01 (0.14) [−0.27, 0.29]
ACE _n	0.001 (0.05) [−0.11, 0.11]	0.01 (0.05) [−0.10, 0.12]	−0.002 (0.06) [−0.12, 0.11]
ACE _a	0.01 (0.07) [−0.12, 0.14]	0.01 (0.06) [−0.12, 0.13]	0.03 (0.07) [−0.12, 0.17]
Conjugate prior number of pseudo subjects, Mean (SE) [95% CI]	PS = 9	PS = 15	PS = 45
ACE _c	0.01 (0.12) [−0.63, 0.91]	0.001 (0.11) [−0.22, 0.22]	0.02 (0.12) [−0.20, 0.23]
ACE _n	−0.002 (0.06) [−0.12, 0.11]	−0.004 (0.06) [−0.12, 0.12]	0.002 (0.06) [−0.11, 0.12]
ACE _a	0.01 (0.07) [−0.13, 0.15]	0.02 (0.07) [−0.12, 0.16]	0.01 (0.07) [−0.13, 0.14]

ξ = inflation factor set by analysis; PS = number of PS in the conjugate prior

to reduce the physician adherence variables at multiple visits to a single indicator of physician adherence, thereby losing information. Others have incorporated principal strata based on multilevel posttreatment variables using principal stratification methods (Mattei and Mealli, 2007). Future work could also use direct marginal modeling for the analysis stage of the multiple imputation inference and could also address the frequentist properties (e.g. finite sample behavior) of the multiple imputation estimator in the CED setting, varying the number of clusters and cluster size, as well as estimator behavior under misspecified models and assumptions.

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGEMENTS

Leslie Taylor, Ph.D., is presently a Mathematical Statistician at the Health Services R&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. Xiao-Hua Zhou, Ph.D., is presently a Core Investigator and Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

FUNDING

The VA Puget Sound Health Care System and Dr. Zhou's VA Research Career Scientist Award (RCD 05-196).

REFERENCES

- ASHTON, C. M., BOZKURT, B., COLUCCI, W. B., KIEFE, C. I., MANN, D. L., MASSIE, B. M., SLAWSKY, M. T., TIERNEY, W. M., WEST, J. A., WHELLAN, D. J. and others (2000). Veterans affairs quality enhancement research initiative in chronic heart failure. *Medical Care* **38**, I26–I37.
- AUDET, A., GREENFIELD, S. AND FIELD, M. (1990). Medical practice guidelines: current activities and future directions. *Annals of Internal Medicine* **30**, 709–714.

- BARNARD, J., DU, J., HILL, J. AND RUBIN, D. B. (1988). A broader template for analyzing broken randomized experiments. *Sociological Methods and Research* **27**, 285–317.
- BARNARD, J., FRANGAKIS, C., HILL, J. AND RUBIN, D. B. (2003). A principal stratification approach to broken randomized experiments: a case study of vouchers in New York city. *Journal of the American Statistical Association* **98**, 299–323.
- FRANGAKIS, C. E. AND 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* **8**, 365–379.
- FRANGAKIS, C. E. AND RUBIN, D. B. (2002). Framework of principal stratification for causal inference in partially controlled studies. *Biometrics* **58**, 21–29.
- FRANGAKIS, C. E., RUBIN, D. B. AND ZHOU, X. H. (2002). Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms. *Biostatistics* **3**, 147–164.
- GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and Bayesian resoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- GRISWOLD, M. E. AND ZEGER, S. L. (2004). On marginalized multilevel models and their computation. Johns Hopkins University, *Department of Biostatistics working papers, working paper 99*. Berkeley, CA: The Berkeley Electronic Press. <http://www.bepress.com/jhubiostat/paper99>.
- HASTINGS, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- HEAGERTY, P. J. AND ZEGER, S. L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science* **75**, 1–26.
- HEITJAN, D. F. (1999). Ignorability and bias in clinical trials. *Statistics in Medicine* **18**, 2421–2434.
- HIRANO, K., RUBIN, D. B. AND ZHOU, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.
- HOLLAND, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–960.
- IMBENS, G. W. AND ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–475.
- IMBENS, G. AND RUBIN, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.
- INSTITUTE OF MEDICINE (1992). *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academy Press.
- LEVY, D. E., O'MALLEY, A. J. AND NORMAND, S.-L. T. (2004). Covariate adjustment in clinical trials with non-ignorable missing data and non-compliance. *Statistics in Medicine* **23**, 2319–2339.
- MATTEI, A. AND MEALLI, F. (2007). Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination. *Biometrics* **63**, 437–446.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. AND TELLER, E. (1953). Equations of state calculations by fast computing machines of input *Journal of Chemical Physics* **21**, 1087–1092.
- NEYMAN, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. *Statistical Science* **5**, 465–480.
- O'MALLEY, A. J. AND NORMAND, S. L. (2005). Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. *Biometrics* **61**, 325–334.
- ROBINS, J. M., GREENLAND, S., AND HU, F. C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association* **94**, 687–700.

- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* **66**, 688–701.
- RUBIN, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6**, 34–58.
- RUBIN, D. B. (1987). Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons.
- RUBIN, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* **5**, 472–480.
- RUBIN, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* **31**, 161–170.
- RUBIN, D. B. (2006). Causal interested through potential outcomes and principal stratification: application to studies with “censoring” due to death. *Statistical Science* **21**, 299–309.
- SHEINER, L. B. AND RUBIN, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* **57**, 6–15.
- SOMMER, A. AND ZEGER S. L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.
- SUBRAMANIAN, U., FIHN, S. D., WEINBERGER, M., PLUE, L., SMITH, F. E., UDRIS, E. M., MCDONELL, M. B., ECKERT, G. J., TEMKIT, M., ZHOU, X. H. and others (2004). A controlled trial of including symptom data in computer-based care suggestions for managing patients with chronic heart failure. *American Journal of Medicine* **116**, 375–384.
- TAYLOR, L. AND ZHOU, X. H. (2009a). Relaxing latent ignorability in the ITT analysis of randomized studies with missing data and noncompliance. *Statistica Sinica* **19**, 749–764.
- TAYLOR, L. AND ZHOU, X. H. (2009b). Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics* **65**, 88–95.
- TEN HAVE, T. R., ELLIOTT, M.R., JOFFE, M., ZANUTTO, E. AND DATTO, C. (2004). Causal models for randomized physician encouragement trials in treatment primary care depression. *Journal of the American Statistical Association* **99**, 16–25.
- ZHANG, J. AND RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. *Journal of Educational and Behavioral Statistics* **28**, 353–368.
- ZHOU, X. H. AND LI, S. (2005). ITT analysis of randomized encouragement design studies with missing data. *Statistics in Medicine* **23**, 1991–2003.

[Received February 5, 2010; revised August 2, 2010; accepted for publication September 13, 2010]