Individual Noncompliance in Clustered Randomized Experiments:
Bayesian Inference and Application to an Encouragement Study
of Advance Directive Forms.

CONSTANTINE E. FRANGAKIS*
Department of Biostatistics, The Johns Hopkins University
Baltimore, MD 21205, U.S.A.

DONALD B. RUBIN
Department of Statistics, Science Center 709, Harvard University
Cambridge, MA 02138, U.S.A.

and

XIAO-HUA ZHOU
Division of Biostatistics, Department of Medicine
Indiana University School of Medicine and
Regenstrief Institute for Health Care Center
Indianapolis, IN 46202, U.S.A.

February 27, 2001

* For correspondence: 615 N. Wolfe St., Baltimore, MD, 21205, tel: (410) 955-3505; fax: (410) 955-0958; email: cfrangak@jhsph.edu

Department of Biostatistics Research Report FRZ00-1, The Johns Hopkins University

Submitted.
Summary

In many randomized experiments comparing a new “target treatment” with a control target treatment, the received treatment does not always agree with assigned treatment - that is, the compliance is not perfect. An obvious example arises when ethical or practical constraints prevent even the randomized assignment of receipt of the new treatment but allow the randomized assignment of the encouragement to receive the new treatment. In fact, in many studies when compliance is not enforced by the experimenter, e.g. with nonblinded assignment, a randomized experiment may be more accurately thought of as a randomized encouragement design. Moreover, often the assignment of encouragement is at the level of clusters (e.g., doctors) where the compliance with the assignment varies across the units (e.g., patients) within clusters. We refer to such situations as “clustered encouragement designs” (CED) and they arise relatively frequently (e.g., Sommer and Zeger, 1991; McDonald, Hiu, and Tierney, 1992; Dexter et al., 1998). Here, we propose methodology for causal inference for the effect of the new target treatment vs. the control target treatment in the CED with all-or-none compliance at the unit level, which generalizes the Bayesian approach of Hirano et al. (2000) in important and surprisingly subtle ways, to account for the clustering, which is necessary for statistical validity. We illustrate our methods using data from a recent study exploring the role of physician consulting in increasing patients’ completion of Advance Directive forms.

Keywords: Advance Directive; Causal Inference; Clustering; Noncompliance; Phenomenological Bayesian Model; Rubin Causal Model.
1. **INTRODUCTION AND PURPOSE.**

1-1. *Motivating studies and data features.*

When evaluating treatment options, direct assignment and enforcement of treatment receipt may not be ethical or feasible. In such cases, it is more realistic to view the design as involving the randomization of encouragement, as opposed to receipt, of the two target treatments, new and standard, were in some designs the encouragement is explicit and no enforcement is even attempted. Commonly, moreover, this encouragement is applied to clusters (e.g., physicians, or villages) of subjects (e.g., patients). An example of such a “clustered-encouragement-design” (CED) was reported by Sommer and Zeger (1991) where investigators randomized villages in Indonesia to offer or not vitamin A supplements to all their infants, but not all infants in the villages assigned to get vitamin A actually received it. Another example of the CED was a study to evaluate a vaccine for influenza, where any potential withholding of the vaccine was considered unethical (McDonald, Hiu, and Tierney, 1992; Hirano, et al., 2000). For this reason, investigators randomized physicians to receive or not receive encouragement to vaccinate their patients, but many patients of the encouraged doctors did not receive a flu shot and some patients of the not-encouraged doctors did receive the shot.

A more recent example of a CED was conducted on Advance Directive (AD) forms (Dexter et al., 1998), which are intended to be completed by patients to allow them to make early decisions about medical treatments at the late stages of life (instructional directives), and designate a representative decision maker (proxy directives) (Wenger et al., 1993). Dexter et al. (1998) randomized physicians to receive or not receive encouragement to discuss AD with patients; the outcome was patient completion of AD, and the original study addressed the effect of encouragement on AD completion (Dexter et al., 1998). For our purpose, however, an equally important substantive research goal is to assess the effect of physicians’ discussion of AD as the new target treatment for potentially increasing patients’ completion rates of the forms relative to the control target treatment of no such discussion (e.g., Miles et al., 1996).
Generally, CED studies share two specific data structure aspects. First, there is frequent noncompliance of individual subjects - not clusters – for the new target treatments within randomized encouragement arms. Second, the distribution of noncompliance and outcomes frequently varies within and between clusters, which are the units of randomization, rather than the individual subjects. We consider CED studies where the compliance for target treatments is by definition or for practical purposes all or none. This type of noncompliance means that there exist, essentially, two subgroups of patients who are not fully identifiable from the data: those who would not change their actual behavior concerning the target treatment no matter what their physician’s assignment – the noncompliers, and those who would comply under both assignments – the compliers (e.g., Imbens and Rubin, 1994; Baker and Lindeman, 1994; Angrist, Imbens, and Rubin, 1996; Baker, 1998; Frangakis and Rubin, 1999). These definitions are local to this particular study and do not suggest compliance or not in other studies.

An intention-to-treat (ITT) analysis is especially appropriate when the randomized intervention is the scientific intervention - the target treatment – of interest. However, the CED uses randomized encouragement only as a surrogate to induce the new target treatment, and ITT analysis is not as appropriate for two reasons. First, the noncompliers arguably do not carry information about the comparison between the target treatments (e.g., biological efficacy or side effects) because, by definition of this group, the randomization does not change receipt of target treatment; for relevant discussion between explanatory and pragmatic comparisons, see Sheiner and Rubin (1996) and Armitage (1998). Second, the noncompliers instead may experience effects of encouragement. For example, in the study on flu shots (McDonald, Hiu, and Tierney, 1992), it may not be ruled out that, for noncomplying physician-patient pairs, the encouragement has triggered physicians to suggest to their patients a number of other precautions against flu, which might not have been taken in the absence of the encouragement. This can induce “encouragement” effects that confound the effect of the vaccine if the noncompliers are included in the ITT analysis.

The second aspect common in CED studies, the clustered structure of units, also has method-
ological implications. Because the assigned encouragement is at the cluster level, assignment is ignorable (Rubin, 1978) only conditionally on the clusters. And, because noncompliance and outcomes can vary both within and between clusters, the interactions between clustering, noncompliance and outcomes need to be addressed.

1.2. **Addressing clustering with noncompliance at the individual level.**

The problem of noncompliance has received increasing attention recently. In particular, it is now generally recognized that the approach of focusing on the compliers, who are not generally fully identifiable from observed data (e.g., Sommer and Zeger, 1991; Baker and Lindeman, 1994; Angrist et al., 1996), is critically different from approaches that use the treatment actually received as if it were randomized, such as “as-treated” or “per-protocol”, and whose bias has been well documented (e.g., Rubin, 1991; Mark and Robins, 1993; Robins and Greenland, 1994; Sheiner and Rubin, 1995). Moreover, implicit assumptions in the standard instrumental variables analyses, such as the a priori exclusion of effects of assignment, have now formal expressions (Angrist et al., 1996), thereby allowing researchers to avoid such exclusion assumptions when they are not plausible. In related work without the exclusion restriction, Robins (1989) derived estimated bounds for treatment effects, Imbens and Rubin (1997a) developed an appropriate Bayesian approach for distinct patient-physician pairs, and, for the latter case, recently, Hirano et al. (2000) have modeled covariate information.

Research on clustered data, on the other hand, has a long history in interconnected literatures including: survey methodology, dating back at least to Cochran (1963); random effects, dating to Hartley and Rao (1967), Harville (1976) and Laird and Ware (1982); estimating equations methods (e.g., Liang and Zeger, 1986); hierarchical Bayesian and Empirical Bayesian methods dating to James and Stein (1961), Efron and Morris (1973), Rubin (1981), etc. Frangakis, Rubin, and Zhou (1998), and Korhonen et al. (2000), have presented early but of limited scope work on clustered randomization with noncompliance.

Here, we investigate the broader combined problem of clustering and individual noncom-
pliance, and thereby propose general methodology for causal inference in studies where these
two data structures are present together, and where structural exclusion restrictions are relaxed.
In the next section we introduce notation and formalize our goal. In Section 3 we discuss mod-
els and methodology: within an abstract phenomenological Bayesian model (Rubin, 1978),
we introduce an appropriate framework for causal inference with clustered data suffering from
noncompliance. We discuss the critical role of clustering and covariates, and propose a flexi-
ble submodel. In Section 4 we illustrate our methods by analyzing data from the study on AD
forms. The final section gives concluding remarks. The appendix gives details on our models.

2. CLUSTERED ENCOURAGEMENT DESIGN

2.1. Setting.

Consider a hospital serving a set of patients, \( i = 1, ..., N \), the \( i \)th patient with physician \( Q_i \),
where \( Q_i = 1, ..., M \leq N \), so that each doctor may serve more than one patient.

In order to compare two target treatments, a new one versus a control one, assume that
the hospital considers two possible actions for each physician: (i) encouraging the physician
to administer the new target treatment, and (ii) no encouragement. In either case, however,
patients within a physician may not all comply with their physician’s assignment. To allow
for this, we adopt the formulation of Angrist et al. (1996) for all-or-none compliance. We
assume that patient \( i \) will either receive the new target treatment, indicated by \( D_i(z) = 1 \), or
the control one, \( D_i(z) = 0 \), when \( Q_i \) is assigned action \( z \), where \( z = 1 \) for encouragement for
the new target treatment and 0 otherwise. Similarly let \( Y_i(z) \) be patient \( i \)’s outcome of interest,
e.g., occurrence/absence of the disease, when physician \( Q_i \) is assigned action \( z \). Covariate
information about patient \( i \) and physician \( Q_i \) is collectively denoted by a \( p \)-dimensional vector
\( \mathbf{X}_i \), which can include an intercept.

Assume, for simplicity, that physicians’ assignments are decided by complete random-
ization, where we let \( Z_i = 1 \) if patient \( i \)’s physician, \( Q_i \), is randomized to encouragement,
and \( Z_i = 0 \) otherwise, and note that, since randomization is in clusters, \( Z_i = Z_{ij} \) whenever
Finally, note that only the values $D_i^{\text{obs}} := D_i(Z_i)$ and $Y_i^{\text{obs}} := Y_i(Z_i)$ are observed; the values under the alternative assignment, $D_i^{\text{mis}} := D_i(1 - Z_i)$ and $Y_i^{\text{mis}} := Y_i(1 - Z_i)$ are missing.

2-2. Compliance subgroups.

Each patient $i$ belongs to one of the following four compliance status groups: $C_i = c$ for a complier, defined by $D_i = (D_i(0), D_i(1)) = (0, 1)$, that is, a patient who would comply with respect to the target treatment under both assignments of physician $Q_i$; $C_i = n$ for a never-taker, that is, a patient who, in this study, would never take the new target treatment no matter the physician’s assignment, so that $D_i = (0, 0)$; $C_i = a$ for an always-taker, one who, in this study, would always take the new target treatment, so that $D_i = (1, 1)$; and $C_i = d$ for a defier, one who would act opposite to the assignment, so that $D_i = (1, 0)$ (e.g., Imbens and Rubin, 1994; Pearl, 1994; Baker and Lindeman, 1994; Angrist et al., 1996; Baker, 1998).

Although the encouragement to use the new target treatment may or may not induce its actual receipt for some patients, we assume that, in this context, the encouragement would not reverse a prior decision to take the new target treatment. This assumption was termed monotonicity by Imbens and Angrist (1994), and allows only for the first three compliance groups, i.e., defiers are not allowed.


By definition, the noncompliers – the never-takers and always-takers, are those for whom different assignment in this experiment would not change their behavior with respect to the target treatments, and therefore those groups are not relevant for comparing the target treatments (e.g., Sommer and Zeger, 1991; Sheiner and Rubin, 1995). Moreover, if the randomized treatment is encouragement for vaccination, then other effects may be present if the encouraged physicians suggest to patients alternative preventive measures, such as recommendations to reduce the patients’ exposure, prescriptions of other medicines, or earlier taking of the vaccine. Such “purely
encouragement” effects are arguably more likely to occur for noncompliers than for compliers, e.g., if the reason some patients are never-takers is precisely the substitution of the target treatment with another when encouraged. Also, discerning “purely encouragement” effects among compliers requires assumptions not verifiable even with knowledge of all compliance statuses in the study. Nevertheless, because any effects of assignment $Z_i$ on outcome $Y$ for noncompliers must be due to sources other than the target treatment $D$, such effects for these patients can be removed simply by focusing analyses on the compliers.

For these reasons we wish to distinguish the noncompliers from the compliers. Using notation consistent with Imbens and Rubin (1997a), we let $l(t) = \{i : C_i = t\}$, the subset of subjects with compliance status $t = c, n,$ or $a$, and let $N_t$ be the number of patients in that group. The standard ITT estimand $\sum_t [Y_i(1) - Y_i(0)]/N$ equals the mixture $\sum_{t \in \{c, a, n\}} N_t \text{ITT}^{(t)}/N$, where

$$\text{ITT}^{(t)} = \frac{1}{N_t} \sum_{i \in l(t)} Y_i(1) - Y_i(0), \quad t = c, a, n,$$  \hspace{1cm} (2.1)

are the average effects of assignment on the compliance status subgroups. In the remainder of our discussion, we focus on estimating the compliance subgroup-specific ITT effects (2.1), and, in particular, the ITT effect on the compliers, $\text{ITT}^{(c)}$. In the next section we investigate the role of clustering and covariates in the CED, and then we bring together these aspects in a model for drawing causal inference.

3. MODELING IN THE CED

3.1. Role of clustering in the phenomenological Bayesian approach.

All potentially observable data can be expressed by the matrix $\mathbf{H} = (\mathbf{D}, \mathbf{Y}, \mathbf{X}, \mathbf{Q}, \mathbf{Z})$, whose $i$th row is the vector $(D_i, Y_i, X_i, Q_i, Z_i)$. Here, $D_i$ denotes the pair of potential receipts $(D_i(0), D_i(1))$ for patient $i$, and $Y_i$ is the pair of potential outcomes $(Y_i(0), Y_i(1))$. Although the quantities $(\mathbf{D}, \mathbf{Y}, \mathbf{X}, \mathbf{Q})$ in $\mathbf{H}$ may be assumed fixed over hypothetical replications of the experiment (as in a permutation-based analysis, Rubin, 1990), we will more generally allow
them to be considered random, with a distribution $\text{pr}(\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q)$. The joint probability distribution of $\mathbf{H}$ that is induced by the distribution $\text{pr}(\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q)$ and the clustered randomization, $\text{pr}(Z|\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q)$, will still be denoted by $\text{pr}$. We assume that the matrix $(\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q)$ contains all the information on observable data and on the design, so that the rows of $\mathbf{H}$ are exchangeable (Rubin, 1978). Following results on exchangeability (de Finetti, 1974), we may essentially write

$$\text{pr}(\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q) = \int \prod_i \text{pr}\{(D_i, Y_i, X_i, Q_i)|\theta\} \text{pr}(\theta) d\theta, \quad (3-1)$$

for some distributions $\text{pr}(\theta)$ and $\text{pr}(D_j, Y_j, X_j, Q_j|\theta)$, where $\theta$ can be thought of as representing the characteristics of a reference population from which the study units, physicians and patients, are drawn.

We stress that, although the clustered randomization on $Q$ is likely to give less precise estimates of effects than a complete randomization at the patient level, the clustered randomization does not affect the joint exchangeability in (3-1). Rather, the clustered randomization is related to the assignment mechanism: because the assignment is at the level of physicians, $Q$, we have that $\text{pr}(Z|\mathbf{D}, \mathbf{Y}, \mathbf{X}, Q) = \text{pr}(Z|Q)$ so that assignment is ignorable (Rubin, 1976, 1978) after but not before conditioning on $Q$. Consequently, inference on the potential outcomes also needs to be conditional on $Q$, which, in this case, is expected to reduce precision because no physician has patients in both assignment arms.

In particular, if the compliance status $C_i$ and the potential outcomes $Y_i(z)$ were known for all $i$ and $z$, then the compliance subgroup-specific effects $\text{ITT}^{(l)}$ could be computed from definition (2-1). Although the values $D_{i}^{\text{obs}} := \{D_{i}^{\text{obs}}\}$ and $Y_{i}^{\text{obs}} := \{Y_{i}^{\text{obs}}\}$ are known, the values $D_{i}^{\text{mis}} := \{D_{i}^{\text{mis}}\}$ and $Y_{i}^{\text{mis}} := \{Y_{i}^{\text{mis}}\}$ are unknown and, under a specific model (3-1), they have a posterior predictive distribution

$$\text{pr}(Y_{i}^{\text{mis}}|D_{i}^{\text{mis}}, H_{i}^{\text{obs}}, \theta) \text{ pr}(D_{i}^{\text{mis}}, \theta|H_{i}^{\text{obs}}), \quad (3-2)$$
where \( \mathbf{H}^{\text{obs}} := (\mathbf{D}^{\text{obs}}, \mathbf{Y}^{\text{obs}}, \mathbf{X}, \mathbf{Q}, \mathbf{Z}) \), the observed data. Then, Bayesian inference on the estimands \( \text{ITT}^{(t)} \) follows from their posterior predictive distributions induced by (3-2).

### 3.2. Role of covariates for compliance-subgroup causal inference.

Because membership in the compliance subgroups is not fully identifiable from observed data, it is important to understand the mechanism through which information on the compliance-subgroup causal effects is recoverable.

For simplicity, assume \( Y_i(z) \) is binary, e.g., 1 for occurrence of disease, 0 otherwise, and let \( \text{ITT}^{(t,\theta)} := \text{pr}(Y_i(1) = 1| C_i = t, \theta) - \text{pr}(Y_i(0) = 1| C_i = t, \theta) \), the average causal effect among people of compliance type \( t = c, n, a \) in the reference population \( \theta \) defined by (3-1) (unconditionally on doctors). The estimand \( \text{ITT}^{(c,\theta)} \) is estimable consistently, as \( N \) grows, under the so-called “exclusion restrictions” (e.g., Sommer and Zeger, 1991; Angrist et al. 1996). However, for the reasons explained in Sec. 1.1, we do not wish to impose a priori these assumptions for the CED. In the absence of exclusion restrictions, and absence of covariates, the effects \( \text{ITT}^{(t,\theta)} \) are not consistently estimable, only bounds thereof are (e.g., Robins, 1989; Manski, 1990; and Balke and Pearl, 1997), under the standard asymptotic sequence of sample size \( N \) growing. Nevertheless, when covariates are available, Frangakis (1999, PhD thesis) shows that the predictive model \( \text{pr}(C_i = t| X_i, \theta) \) of compliance status from the covariates is estimable with no exclusion restrictions, and that for the same effects \( \text{ITT}^{(t,\theta)} \), asymptotic bounds are narrower if the covariates are used than if they are not used, so that \( \text{ITT}^{(t,\theta)} \) are consistently estimable under an asymptotic theory where, in addition to \( N \), the number of covariates grows in an appropriate sequence.

In practice, the above discussion does not change the way Bayesian inference is drawn, but helps critically in understanding the role of covariates in recovering information. Under a Bayesian model for (3-1), inference is based directly on the posterior predictive distribution of the target estimands \( \text{ITT}^{(t)} \) induced by (3-2), where (i) in the absence of covariates, the spread of the posterior distribution will reflect the relatively large uncertainty in predicting compliance
status, whereas (ii) when a covariate that predicts compliance status is modeled, the spread of the posterior distribution of the estimands will be narrower.

In the remaining part of this paper, we describe a specific Bayesian sub-model of (3.1) and apply it to the study of AD introduced in Section 1. Inference using estimated covariate-adjusted asymptotic bounds and comparisons with Bayesian inference will be discussed in detail in a subsequent article.

3-3. Specific model.

We model the joint distribution \( \text{pr}(D_i, Y_i, X_i, Q_i, \theta) \) of (3.1) in conditional distributions. First, because we focus on the finite study estimands ITT\(^{(c)}\) and, because \( X_i \) and \( Q_i \) are known for all \( i \), for our model we take the marginal distribution of \( X_i \) and \( Q_i \) to be the observed distribution. Conditionally on \( X, Q \), for the other model components we assume the following structure.

Compliance Status.

We model the compliance status with two probit submodels,

\[
\text{pr}(C_i = n|X_i, Q_i, \theta) = 1 - \Phi(X_i'\alpha_{(c,1)} + W_i'b_{Q_i}^{(c,1)}),
\]

\[
\text{pr}(C_i = c|X_i, Q_i, \theta) = \{1 - \text{pr}(C_i = n|X_i, Q_i, \theta)\}\{1 - \Phi(X_i'\alpha_{(c,2)} + W_i'b_{Q_i}^{(c,2)})\},
\]

and \( \text{pr}(C_i = a|X_i, Q_i, \theta) = 1 - \text{pr}(C_i = c|X_i, Q_i, \theta) - \text{pr}(C_i = n|X_i, Q_i, \theta) \). In these expressions, \( \alpha_{(c,1)}, \alpha_{(c,2)} \) are \( p \times 1 \) parameter vectors that model the association between compliance status and covariates, which can be critical for increasing precision. The vector \( W_i \) is an \( r \times 1 \) subset of \( X_i \) that can include the intercept and characteristics that vary across patients clustered within physician \( Q_i \). The corresponding parameters \( b_{Q_i}^{(c,1)}, b_{Q_i}^{(c,2)} \) are \( r \times 1 \) vectors specific to physician \( Q_i \), and these model the association between compliance status and physicians, to address the assignment mechanism. The function \( \Phi() \) is the standard normal cumulative distribution.

To facilitate computations later, we augment the above model with two latent variables for
each person, \((V_i, \text{ and } U_i)\), defined via the relations

\[
\begin{align*}
C_i &= n \text{ if } C_i(n)^* = X_i'\alpha^{(c,1)} + W_i'\beta^{(c,1)} + V_i \leq 0, \\
C_i &= c \text{ if } C_i(n)^* > 0 \text{ and } C_i(c)^* = X_i'\alpha^{(c,2)} + W_i'\beta^{(c,2)} + U_i \leq 0,
\end{align*}
\]

and where \(V_i \sim N(0, 1)\) and \(U_i \sim N(0, 1)\) independently.

**Potential Outcomes.** For our application in Sec. 4 we have binray potential outcomes so we posit the probit model,

\[
\text{pr}(Y_i(z) = 1 | C_i = t, X_i, Q_i, \theta) = \Phi\{f^{(1)}(X_i, t, z)\alpha^{(y)} + f^{(2)}(W_i, t)\beta^{(y)} \}
\]

for \(t = c, n, a\) and \(z = 0, 1\). Here, \(f^{(1)}, f^{(2)}\) are link vector functions of dimensions \(p_f, r_f\), respectively; \(\alpha^{(y)}\) is a \(p_f \times 1\) parameter vector; and \(\beta^{(y)}\) are \(r_f \times 1\) parameters specific to physician \(Q_i\). This formulation assumes, for parsimony, that, given \(f^{(1)}(X_i, C_i, z)\alpha^{(y)}\), the degree of additional heterogeneity of patients’ outcomes across physician, expressed by \(f^{(2)}(W_i, C_i)\beta^{(y)}\), is similar across randomized arms. Also, we assume that, given the covariates, physician indicators, compliance statuses, and parameters, the two potential outcomes \(Y_i(1)\) and \(Y_i(0)\) are independent. Deviations from this assumptions can be allowed, but would not enter as a parameter in the likelihood, and, as in related settings, would have diminishing inferential impact when the number of subjects averaged in the estimand is large enough (Imbens and Rubin, 1997; Hirano et al., 2000).

As with the model for compliance, to facilitate computations, we could augment the model of outcome in the sense

\[
Y_i(z) = 1 \text{ if } Y_i(z)^* = f^{(1)}(X_i, C_i, z)\alpha^{(y)} + f^{(2)}(W_i, C_i)\beta^{(y)} + S_i(z) \geq 0
\]

where, for each \(z, S_i(z) \sim N(0, 1)\) independently of \(U_i, V_i\) for \(i = 1, ..., N\). We emphasize that this independence is not related to independence (or dependence) between outcome and com-
pliance status. Furthermore, our assumptions of conditional independence between \( Y_t(0) \) and \( Y_t(1) \) at this stage would imply independence between \( S_t(0) \) and \( S_t(1) \). For our computations, we need only consider the augmentation \( S_t = S_t(Z_t) \) (see Appendix A, item 4).

**Parameters.**

The parameter \( \theta \) includes \( \{\alpha^{(C,1)}; \alpha^{(C,2)}; \alpha^{(Y)}; b^{(C,1)}_q, b^{(C,2)}_q, b^{(Y)}_q, q = 1, ..., M; D^{(b)}\} \), where \( D^{(b)} \) is a variance-covariance matrix parameter yet to be introduced. We choose the prior distributions for \( \theta \) to be proper but diffuse, in order to ensure proper posterior distributions and relatively fast convergence of the fitting algorithms, but, at the same time, to be relatively noninformative for our application (Sec. 4). Given the doctor-specific parameters, we posit prior distributions

\[
\alpha^{(C,1)} \sim N(\alpha_0^{(C,1)}, I\xi), \quad \alpha^{(C,2)} \sim N(0, I\xi), \quad \alpha^{(Y)} \sim N(0, I\xi),
\]

(3.5)

independently. Here, \( I \) is the identity matrix, and \( \xi \) is an inflating factor, which is set by the analyst. The component of \( \alpha_0^{(C,1)} \) that corresponds to the intercept is set to \( -\Phi(\frac{1}{3})\sqrt{\xi} \) and the remaining components are set to 0 to represent a prior proportion of approximately 33% for each of the three compliance groups.

The physician-specific parameters are assumed random with

\[
(b^{(C,1)}_q, b^{(C,2)}_q, b^{(Y)}_q) | D^{(b)} \sim N(0, D^{(b)}),
\]

(3.6)

independently across physicians \( q = 1, ..., M \). In our application of Sec. 4 we had a relatively low proportion of always-takers, so we constrained \( D^{(b)} \) so that \( (b^{(Y)}_q, b^{(C,1)}_q) \) and \( b^{(C,2)}_q \) were a priori independent. For the other components of \( D^{(b)} \), we assume that the inverses of \( D^{(1)} := \text{var}(b^{(Y)}_q, b^{(C,1)}_q) \) and \( D^{(2)} := \text{var}(b^{(C,2)}_q) \) a priori have Wishart distributions with scale matrices \( (\beta^{(1)} R^{(1)})^{-1} \) and \( (\beta^{(2)} R^{(2)})^{-1} \), respectively, and degrees of freedom, \( \beta^{(1)} \) and \( \beta^{(2)} \) respectively, equal to the dimensions of \( (b^{(C,1)}_q, b^{(Y)}_q) \) and \( b^{(C,2)}_q \) (in order to introduce relatively little prior information, e.g., Wakefield et al., 1994; for \( R^{(1)} \) and \( R^{(2)} \) see appendix), although it
would also be interesting to study more informative priors.

Using model (3.3)-(3.6), inference on the estimands of interest, \( \text{ITT}^{(t)} \), \( t = c, a, r_1 \), in the finite study population is based on (3.2). The appendix outlines an algorithm for simulating these distributions for our models.

4. APPLICATION TO ADVANCE DIRECTIVES

4.1. Procedures.

We now return to the study of AD forms introduced in Sec. 1.1. AD forms are designed to increase patient’s autonomy, and although enjoy support by ethicists, and physicians (Hughes and Singer, 1992), very few patients complete them in practice, and very few physicians discuss the role of these forms with their patients. It has been hypothesized that if physicians briefly discussed the role of AD forms with their patients, this would cause completion rates to substantially increase (Miles et al. 1996). The discussion effect is very important because, if shown large, could help convince physicians to spend the brief time needed to discuss AD with even more of their eligible patients. So, our goal is to address this hypothesis from a CED taking into account the different sources of variation. The problem of more flexible alternative designs in this application is studied by Frangakis and Baker (2001).

The data we use are a subset from the study on AD forms by Dexter et al. (1998), who analyzed the data by ITT analyses. In that study the researchers randomly divided eligible physicians of an urban hospital into four groups: one group routinely received computer reminders to discuss instructional directives with their patients; another group received reminders on proxy directives; a third group received both reminders, and a fourth group received no reminders. Here, the subset of data available is from the control group, denoted by \( (Z = 0) \), and from the group receiving reminders for both AD forms, denoted by \( (Z = 1) \), where only patients eligible for AD discussion and completion were included. To use these data in our framework, let \( D_{i}^{obs} \) be the indicator for actual discussion of AD, equal to 1 if patient \( i \)'s doctor discusses any of the two AD forms with patient \( i \) (the new target treatment), 0 otherwise (the control target
treatment), and $Y^{\text{obs}}_t$ be the indicator for completion of AD, equal to 1 if the patient completes any AD form, 0 otherwise. Patient age data are also of interest because age has been previously shown to be associated with discussions of AD forms (Duffield and Podzamsky, 1996; Boyd et al., 1996; Hakim et al., 1996). Moreover, because in a preliminary analysis, we found none of the available covariates other than age to be useful in predicting compliance status, here we use only it. Table 1 gives some basic characteristics of our data that can be summarized easily using theory for finite population cluster sampling (Cochran, 1963).

Table 1 about here

Table 1 shows that 3% of patients completed AD forms under control ($Z = 0$) versus 14% when their physicians were encouraged to discuss AD ($Z = 1$). However, approximately 74% of the physician-patient pairs did not discuss the forms when encouraged, and so, under Sec. 2.2, must be never-discussants in the sense that they would not discuss the forms whether encouraged or not in this study. Analogously, approximately 5% of the pairs are always-discussants. The estimated remaining 22% of physician-patient who are neither always-discussants nor never-discussants are discussion-compliers, in the sense that they would discuss AD if and only if encouraged. Therefore, although the modest 11% ITT effect of encouragement on completion rates may suggest to physicians that discussing AD forms with their patients has no practical effect, the majority of physician-patient counted in that estimate are not relevant to the effect of AD discussion on AD completion.

To address this, we argue that the evidence for the effect of AD discussion on AD completion needs to be sought among discussion compliers. Also, although it is not known whether physicians or patients initiated the discussions, focusing on discussion-complier pairs almost ensures that the (new) discussions under encouragement are initiated by physicians, because they were the unit of intervention. Therefore, we focus on estimating the effect of encouragement on completion among the subset of discussion-compliers, ITT$(e)$, which, here, we call the effect of discussion on form completion. To estimate ITT$(e)$, we use two procedures based on models of the form (3.3)-(3.6).
The first model-based procedure is based on the specifications (3-3)-(3-6) with no covariates. For this model, $X_i$ and $W_i$ are 1. The outcome link function $f^{(1)}$ saturates the patterns of compliance status by treatment arm: 

$$f^{(1)}(X_i, C_i, z) = [I(C_i = n) * z, I(C_i = n) * (1 - z), I(C_i = c) * z, I(C_i = c) * (1 - z), I(C_i = a) * z, I(C_i = a) * (1 - z)].$$

We take the vector function $f^{(2)}(W_i, C_i)$ to be $[I(C_i = n), I(C_i = c), I(C_i = a)]$. The hyperparameter $\xi$ in (3-5) is set here to 5, although other values were also tried (see also Sec. 4.3).

The second procedure also uses patients’ age in the model components (3-3)-(3-6). Specifically, for this model, $X_i$ and $W_i$ are set to $[1, \text{age}_i]$, where $\text{age}_i$ is the patient’s age in standardized log-log scale. The link function $f^{(1)}$ now fits separate intercepts and slopes on $\text{age}_i$ for each of the six combinations of compliance status $C_i$ crossed by assignment arm $z$. The link function $f^{(2)}$ fits separate intercepts and slopes on $\text{age}_i$ for each compliance status. The values of the hyperparameters are as in the model without age. For further details on the fitting methods see the appendix.

4.2. Results.

Table 2 about here

In Table 2, we report estimates of the estimands of compliance subgroup-specific AD completion percentages under control, under encouragement, and the between-arm (encouragement - control) difference ITT\textsuperscript{(l)}, $t = c, n, a$ using the two model-based procedures. For each model, we obtain the posterior predictive distributions of these estimands, as induced by (3-2). The table summarizes these distributions by means (within treatment arms, the posterior predictive means are the posterior predictive probabilities of AD completion averaged over $X_i, Q_i$ and $\theta$ from its posterior), standard deviations, and 95% intervals (2.5% and 97.5% quantiles). The results closely follow the two main points on the role of clustering and covariates discussed, respectively, in Sections 3.1 and 3.2.

In particular, the model that accounts for clustering but does not model age gives, mostly, unhelpfully broad answers. An exception is inference for the never-discussants. This is because
the observed completion rate for the non-discussants under control arm (Table 1), which is, generally, the mixture of completion rates for complier-discussants and never-discussants in that arm, is null here and, so, implies null completion rate for both complier-discussants and never-discussants in the random half of the patients who were assigned in that arm. However, most of the other 95% intervals, including for the effect on complier-discussants, are too wide for practical use or interpretation with this model.

In contrast, the model that uses age in both the compliance and the outcome components provides 95% posterior intervals that are quite usefully narrower than those of the model without age. In particular, among compliers, the effect of assignment on completion rates has a posterior mean of 62% and is most likely at least 34% (2.5% posterior quantile). These estimates are also substantially higher than those reported by the ITT analyses in Table 1, and reflect in a principled way the uncertainty from the different sources of missing information.

The other results are generally not surprising, except perhaps the negative estimates for the effect of encouragement among always-discussants. Nevertheless, because the null is well within the posterior interval in both models, this result is consistent with random fluctuation.

Table 3 about here

It is relevant to check the extent to which the increased precision of the second model-based procedure can be attributed to information in the data, including the ability of age to predict compliance status, or information supplied by the prior distribution. In Table 3 we report the probabilities of completion rates under control, encouragement and their difference as induced solely by the prior distributions for each model. For example, to get a draw $Y(0)$ from the prior distribution for compliers, we (i) draw a subject’s $X_i, Q_i$ from $\text{pr}(X_i, Q_i)$ (by assumption here, the observed distribution), (ii) draw $\theta$ from the defining model $\text{pr}(\theta|X_i, Q_i) = \text{pr}(\theta)$, (iii) draw compliance status $C_i$ from the model $\text{pr}(C_i = t|X_i, Q_i, \theta), t = c, n, a$ of (3·3), and (iv) if the status is “complier”, we calculate $\text{pr}(Y_i(0) = 1|C_i = c, X_i, Q_i, \theta)$ from (3·4) and, with this probability, draw a Bernoulli outcome $Y(0)$. The distributions for the other entries are derived analogously. None of these distributions appears particularly informative, suggesting
that the increased precision in the model with age in Table 2 is not particularly influenced by
the chosen prior distributions.

Moreover, the posterior distributions of the model parameters showed evidence that the
probability of being a never-taker decreased with age (mean, [2.5%, 97.5%] quantiles for \( \alpha_{\text{age}}^{(C,1)} \): 0.052 [0.025, 0.094]), and most of the shift was to being a complier (\( \alpha_{\text{age}}^{(C,2)} \): -1.293 [-2.275, -0.122]). We also calculated the posterior predictive distribution of the probability that the next
patient seen has each compliance status as a function of that patient’s age. Means and standard
deviations from this distribution conditionally on age are displayed in Figure 1. Each draw from
this distribution is obtained as the defining probability of compliance status in (3-3), where \( \alpha^{(C,1)} \)
and \( \alpha^{(C,2)} \) are drawn from their posterior distribution, and \( b_q^{(C,1)} \) and \( b_q^{(C,2)} \) are drawn from their
defining model (3-6) after having drawn \( D(\theta) \) from its posterior distribution.

These results combined are in agreement with the expectation that covariates that predict
compliance status are critical for obtaining useful inference in problems where the compliance
status membership is not fully observed in the data.

Figure 1 about here

Based on these results, the estimates obtained from the model with age are expected to
be more appropriate for the compliance subgroup-specific effects than the model without age.
Taking the effect of encouragement on the discussion-compliers to be the relevant estimand
for the effect of AD discussions on AD completion, these results give support to, and consid-
erably strengthen the hypothesis that physician-discussion can substantially increase patient-
completion of AD forms.

4.3. Other models.

More than twenty other models of the form (3-3)-(3-6) were fit, were we varied the link functions
\( f^{(1)} \) and \( f^{(2)} \), the functional forms for age, and the inflating factor \( \xi \) in (3-5). In addition, in a
preliminary effort in this work (Frangakis, Rubin, and Zhou, 1998), we had also tried logistic
mixed effects analogues to the probit models reported here. Those models gave results mostly
similar to the ones presented in Table 2 here (see, for example, Table 2 of Frangakis, Rubin, and Zhou, 1998), although with varying performance in the criteria of (i) degree in which the prior distributions influenced the results in the sense of the measures in Table 3; and (ii) convergence diagnostics. For example, in contrast to the models in Sec. 3.3, the simulation stage for the logistic mixed effects model, after incorporation of Metropolis-Hastings adjustments to address lack of conjugacy, did not pass the convergence diagnostics of Gelman and Rubin (1992) (see also appendix) within a satisfactory time for this problem. Among the models we tried, the two reported in the previous section were the most acceptable with respect to these two criteria.

5. Remarks

We described methodology for causal inference in studies with randomization in clusters but noncompliance at the individual level. The proposed method improves upon current procedures, which face limitations with respect to either validity or precision in estimation.

Our methods assume all-or-none observed compliance. For situations where observed compliance is continuous or multilevel, an approach that would discretize compliance to two (or few) levels can still be practically useful, as is often common practice with continuous variables (e.g., age simplified to young/old) where appropriate. Alternatively, direct modelling of the multiple compliance statuses could be done, for example, by aliasing appropriate parameters while keeping a model that would contain the null. It would then be increasingly relevant to explore the plausibility of additional assumptions that would be required to analyze such more structure data, and compare to alternative methods. However, we believe it is precisely the emphasis of our approach to the existing latent groups in these problems, that (a) allows the researcher to input (or not) in the analysis scientific assumptions from among a larger class; and (b) makes the role of various assumptions explicit.

Our procedure is designed to be Bayesian for the input prior. In general, therefore, a posterior interval (PI) does not automatically share the property of a confidence interval (CI) just as a CI does not automatically share the property of the PI. It would be interesting to study cali-
bation of Bayesian and more traditional frequency properties in this problem, for example, by asymptotes that allow information from covariates to grow with samples size, as mentioned in Sec. 3.2, for approximations, or by mixing permutation distributions with the Bayesian model (e.g., Rubin, 1998) for simulation-based solutions.

A MODEL FITTING

Computation of the posterior distribution (3-2) of the missing compliance statuses, say, $C_{\text{mis}}$, missing potential outcomes $Y_{\text{mis}}$, and parameters $\theta$ were based on simulations from a suitable Gibbs sampler (Geman and Geman, 1984). The Gibbs sampler we used draws, in this order: the missing compliance statuses $C_{\text{mis}}$; the missing potential outcomes $Y_{\text{mis}}$; the latent variables $C_i(n)^* \text{ and } C_i(c)^*$ for the current set of never-takers compliers and always-takers; the latent variables $Y_i^* \equiv Y_i(Z_i)^*$ for the outcome model; the parameters $\alpha^{(C)}_1, \alpha^{(C)}_2$ for the compliance model; the outcome model parameters $\alpha^{(Y)}$; the cluster-specific parameters $b_q^{(Y)}, b_q^{(C)}_1, b_q^{(C)}_2$, for $q = 1, \ldots, M$; and the variance matrices $D^{(1)}, D^{(2)}$. For all steps, drawing is done cyclically and each step conditions on all other unknowns, with the following exceptions: the first step must exclude $C_i(n)^*$ and $C_i(c)^*$ from the conditioning in order for the Gibbs sampler to converge to the posterior distribution; also, at this step, the conditional distribution on $Y_i^{\text{obs}}$ is relatively easy to simulate from, and, so, replaces the conditional distribution on $Y_i^*$ for algorithmic efficiency; and, the potential outcomes $Y_{\text{mis}}$, drawn at step 2 to calculate the estimands (2-1), are not included in any other conditioning, for algorithmic efficiency. The distributions involved in the Gibbs sampler are as follows.

1. Any compliance status that is missing, is drawn at this step from $\Pr(C_i|Y_i^{\text{obs}}, D_i^{\text{obs}}, X_j, Q_i, Z_i, \theta)$. This distribution is obtained from the joint distribution $\Pr(C_i, Y_i^{\text{obs}}, D_i^{\text{obs}}|X_j, Q_i, Z_i, \theta)$. For example, a subject with $Z_i = D_i^{\text{obs}} = 0$ can be either a complier or a never-taker, and the conditional Bernoulli distribution of $C_i$ is proportional to

$$\{l(c, Z_i, X_j, Q_i, Y_i^{\text{obs}}, \theta)\}^{I(C_i = c)} \{l(n, Z_i, X_j, Q_i, Y_i^{\text{obs}}, \theta)\}^{I(C_i = n)}$$
where we define \( l(t_0, z_0, x_0, q_0, y_0, \theta) \) to be

\[
\text{pr}(C_i = t_0 | \underline{X}_i = x_0, Q_i = q_0, \theta) \cdot \text{pr}(Y_i(z) = y_0 | C_i = t_0, \underline{X}_i = x_0, Q_i = q_0, \theta).
\]

Therefore, the conditional probability of the subject being a complier at this step is

\[
l(c, Z_i, \underline{X}_i, Q_i, Y_i^{obs}, \theta) | \{ l(c, Z_i, \underline{X}_i, Q_i, Y_i^{obs}, \theta) + l(n, Z_i, \underline{X}_i, Q_i, Y_i^{obs}, \theta) \}^{-1}.
\]

The drawing of \( C_i \) for subjects with \( Z_i = D_i^{obs} = 1 \) is done in a similar way.

2. The missing potential outcome, \( Y_i^{mis} = Y_i(1 - Z_i) \), of each person is drawn from the Bernoulli distribution with probability

\[
\text{pr}(Y_i(z') = 1 | C_i = t, \underline{X}_i, Q_i, Y_i(1 - z'), \theta) = \text{pr}(Y_i(z') = 1 | C_i = t, \underline{X}_i, Q_i, \theta),
\]

that is, the defining model (3-4), where \( z' \) is set to \( 1 - Z_i \).

3. The drawing of \( C_i(n)^* \) is from \( \text{pr}(C_i(n)^* | \underline{X}_i, Q_i, C_i, \theta) \). This distribution is the same as the defining model \( \text{pr}(C_i(n)^* | \underline{X}_i, Q_i, \theta) \) but truncated either to the left or to the right of zero depending on \( C_i \). The drawing of the truncated normal is done using its inverse distribution function, which is readily calculable. For subjects that, in the previous cycle of the algorithm, had been imputed as always-takers or compliers, the drawing of \( C_i(c)^* \) is done in a similar way.

4. The drawing of \( Y_i^* \) is from \( \text{pr}(Y_i(z)^* | \underline{X}_i, Q_i, Y_i^{obs}, \theta) \), where \( z \) is set to \( Z_i \). This distribution is the same as the defining model \( \text{pr}(Y_i(z)^* | \underline{X}_i, Q_i, \theta) \) except that it is truncated to the right or left of zero depending on \( Y_i^{obs} \). The drawing is as with the compliance latent normals.

5. The drawing of the coefficients \( \alpha^{(C,1)} \) is from \( \text{pr}(\alpha^{(C,1)} | \{ \text{ all } C_i(n)^*, \underline{X}_i, Q_i, b_{Q_i}^{(C,1)} \}) \). This distribution is a Bayesian linear regression based on the defining likelihood and prior with offsets \( \underline{W}_i b_{Q_i}^{(C,1)} \) known at this step. The drawing of the coefficients \( \alpha^{(C,2)} \) is from \( \text{pr}(\alpha^{(C,2)} | \{ \text{ all } C_i(c)^*, \underline{X}_i, Q_i, b_{Q_i}^{(C,2)} : C_i = a \text{ or } c \}) \), and the drawing of the coefficient \( \alpha^{(Y)} \) is from \( \text{pr}(\alpha^{(Y)} | \{ \text{ all } Y_i^*, \underline{X}_i, Q_i, Z_i, C_i, b_{Q_i}^{(Y)} \}) \), both of which are Bayesian linear regressions with offsets, respectively, \( \underline{W}_i b_{Q_i}^{(C,2)} \) and \( f^{(2)}(\underline{W}_i, C_i) b_{Q_i}^{(Y)} \).
6. Cluster-specific parameters $b_q^{(y)}$ are independently drawn for each cluster $q = 1, \ldots, M$ from $\Pr(b_q^{(y)} | \{ \text{all } Y_i^*, X_i, Z_i, C_i : Q_i = q \}, \alpha^{(y)})$ which is a Bayesian regression with offsets $f^{(1)}(X_i', C_i, Z_i) \alpha^{(y)}$. Then, the prior mean and variance matrix for $\{b_q^{(c,1)}\}$ are adjusted to the conditional prior mean and variance matrix given the drawn values of $b_q^{(y)}, q = 1, \ldots, M$ and used as priors in the Bayesian regression $\Pr(b_q^{(c,1)} | \{ \text{all } C_i(n)^*, X_i : Q_i = q \}, \alpha^{(c,1)})$, with offsets $X'_i \alpha^{(c,1)}$, to draw $b_q^{(c,1)}$. The parameters $b_q^{(c,2)}$ are drawn analogously, except that the prior mean and variance matrix are not adjusted for the drawing of $b_q^{(y)}, b_q^{(c,1)}$ because of the assumed prior independence.

7. The drawing of $\{D^{(1)}\}^{-1}$ is from the Wishart distribution with $M + \beta^{(1)}$ degrees of freedom and scale matrix $[\sum_q (b_q^{(y)}, b_q^{(c,1)})(b_q^{(y)}, b_q^{(c,1)})' + \beta^{(1)} R^{(1)}]^{-1}$ (Gelman et al., 1995). The drawing of $\{D^{(2)}\}^{-1}$ is from the Wishart distribution with $M + \beta^{(2)}$ degrees of freedom and scale matrix $[\sum_q b_q^{(c,2)} b_q^{(c,2)' + \beta^{(2)} R^{(2)}]}^{-1}$.

For each model, three chains were run. Independently among chains and across subjects, any unknown compliance status was initiated as Bernoulli draw, conditionally on assignment arm, and with probabilities of compliance status obtained from the sampling point estimates derived from Table 1. Subsequently, parameter estimates were initialized based on generalized linear models estimates: for $\alpha^{(c,1)}$ and $\alpha^{(y)}$, based on all subjects; for $\alpha^{(c,2)}$ based on the initialized set of compliers and always-takers; and for the physician-specific parameters, based on physicians with corresponding full rank design matrices. The initialized physician-specific parameters were used to set the values of $R^{(1)}$ and $R^{(1)}$, to represent, respectively, preliminary estimates of $D^{(1)}$ and $D^{(2)}$ (e.g., Wakefield et al., 1994). The matrices $D^{(1)}$ and $D^{(2)}$ were initialized to the values of $R^{(1)}$ and $R^{(1)}$, respectively. Each chain was run for 25000 iterations. At 12500 iterations, and based on the three chains for each model, the potential scale reduction statistic (Gelman and Rubin, 1992) was computed for ITT(t), $t = c, n, a$ giving 1.00, 1.02, 1.02 respectively, suggesting no evidence against convergence. Inference for each model of Table 2 is based on the remaining 37500 iterations, combining the three chains.
REFERENCES


Biometrika 86, 365–379.
Berkeley, 361–379.


Table 1: Patient characteristics in our sample from the study of Advance Directives. Estimates are based on sampling-theory ratio estimation for cluster sampling of finite population (Cochran, 1963, p. 30).

<table>
<thead>
<tr>
<th></th>
<th>Control assignment</th>
<th>Encouragement assignment</th>
<th>Difference between assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctors (no.)</td>
<td>26</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Patients (no.)</td>
<td>158</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>64.6 (0.6)</td>
<td>64.1 (0.4)</td>
<td>−0.5 (1.0) [−0.5]</td>
</tr>
<tr>
<td>Discussing AD, %</td>
<td>5.1 (1.2)</td>
<td>25.7 (3.9)</td>
<td>21.6 (5.8) [3.7]</td>
</tr>
<tr>
<td>Completing AD, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>among discussants</td>
<td>62.5 (8.6)</td>
<td>51.1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>among non-discussants</td>
<td>0.0 —</td>
<td>1.5 (0.8)</td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>3.2 (0.9)</td>
<td>14.3 (3.2)</td>
<td>11.1 (4.6) [2.4]</td>
</tr>
</tbody>
</table>

Estimates are mean (se). Numbers in brackets are ratios of estimated mean over standard error. Difference is encouragement minus control assignment.
Table 2: Completion rates (%) of Advance Directives for compliance subgroups and assignment arms: (i) the model of Section 3.3 without covariates, and (ii) the model including age in both the compliance status and potential outcomes component. Reported results from the models are means (standard deviations) and 95% intervals [2.5% and 97.5% quantiles] of the corresponding posterior predictive distributions.

<table>
<thead>
<tr>
<th>AD Completion, %</th>
<th>Discussion-Compliers</th>
<th>Never-Discussants</th>
<th>Always-Discussants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>3.1 (6.4)</td>
<td>0.2 (0.4)</td>
<td>62.0 (14.4)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 21.6]</td>
<td>[0.0, 1.6]</td>
<td>[31.7, 86.7]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>3.3 (4.0)</td>
<td>0.2 (0.4)</td>
<td>46.7 (12.2)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 13.7]</td>
<td>[0.0, 1.2]</td>
<td>[26.9, 71.6]</td>
</tr>
<tr>
<td>Encouragement assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>47.9 (21.3)</td>
<td>1.9 (1.2)</td>
<td>47.9 (31.1)</td>
</tr>
<tr>
<td></td>
<td>[4.5, 89.7]</td>
<td>[0.8, 5.1]</td>
<td>[0.0, 100.0]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>65.3 (9.9)</td>
<td>2.1 (1.2)</td>
<td>34.8 (9.7)</td>
</tr>
<tr>
<td></td>
<td>[41.9, 80.9]</td>
<td>[0.8, 5.2]</td>
<td>[17.2, 54.2]</td>
</tr>
<tr>
<td>Difference between assignments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>44.7 (22.4)</td>
<td>1.7 (1.4)</td>
<td>-14.1 (33.9)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 86.2]</td>
<td>[-0.4, 5.1]</td>
<td>[-68.9, 46.7]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>62.0 (11.2)</td>
<td>1.9 (1.3)</td>
<td>-11.8 (14.4)</td>
</tr>
<tr>
<td></td>
<td>[34.7, 79.5]</td>
<td>[0.0, 4.9]</td>
<td>[-42.1, 14.7]</td>
</tr>
</tbody>
</table>
Table 3: Prior distributions. Induced probabilities of completion (%) of Advance Directives for compliance subgroups and assignment arms using: (i) the prior distribution for the model of Section 3.3 without covariates, and (ii) the prior distribution of the model that included age. Reported results are means (standard deviations) and 95% intervals [2.5% and 97.5% quantiles] of the corresponding prior distributions.

<table>
<thead>
<tr>
<th>AD Completion, %</th>
<th>Discussion-Compliers</th>
<th>Never-Discussants</th>
<th>Always-Discussants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>49.0 (50.0)</td>
<td>48.4 (50.0)</td>
<td>49.8 (50.0)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>50.3 (48.4)</td>
<td>48.7 (48.1)</td>
<td>53.5 (48.7)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
</tr>
<tr>
<td>Encouragement assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>49.8 (50.0)</td>
<td>49.2 (50.0)</td>
<td>49.2 (50.0)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>50.2 (48.0)</td>
<td>49.0 (48.1)</td>
<td>52.5 (48.6)</td>
</tr>
<tr>
<td></td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
<td>[0.0, 100.0]</td>
</tr>
<tr>
<td>Difference between assignments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model (w/o age)</td>
<td>0.9 (66.6)</td>
<td>0.8 (68.7)</td>
<td>-0.6 (63.9)</td>
</tr>
<tr>
<td></td>
<td>[-100.0, 100.0]</td>
<td>[-100.0, 100.0]</td>
<td>[-100.0, 100.0]</td>
</tr>
<tr>
<td>Model (w/ age)</td>
<td>-0.1 (59.8)</td>
<td>0.8 (65.0)</td>
<td>-1.1 (52.2)</td>
</tr>
<tr>
<td></td>
<td>[-100.0, 100.0]</td>
<td>[-100.0, 100.0]</td>
<td>[-100.0, 100.0]</td>
</tr>
</tbody>
</table>
Figure 1. Age and compliance status in the study on Advance Directive forms. Posterior predictive distribution of probabilities of being a discussion-complier (mean is height of lower solid curve), a never-discussant (mean is height between solid curves), and an always-discussant (mean is height between upper solid curve and 1.0). The two dotted lines around the lower (upper) solid curve are ±1 posterior standard deviation of the probability of complier (always-discussant) (see also Sec. 4.2).