In Biometrics, 2001, 57, 899–908.

Compliance sub-sampling designs for comparative research: estimation and optimal planning.

Constantine E. Frangakis
Department of Biostatistics, The Johns Hopkins University
615 N. Wolfe St., E2006, Baltimore, MD, 21205, USA.
www.jhsph.edu/Departments/Biostats/research/frangakis.html
cfrangak@jhsph.edu

Stuart G. Baker
Biometry Research Group, National Cancer Institute
EPN 344, 6130 Executive Blvd. MSC 7354, Bethesda, MD 20892-7354, USA.
http://dcp.nci.nih.gov/bb/baker.html
email: sb16i@nih.gov.

July 5, 2002

SUMMARY. For studies with treatment noncompliance, analyses have been developed recently to better estimate treatment efficacy. However, the advantage and cost of measuring compliance data have implications on the study design that have not been as systematically explored. In order to estimate better treatment efficacy with lower cost, we propose a new class of "compliance sub-sampling" (CSS) designs where, after subjects are assigned treatment, compliance behavior is measured for only subgroups of subjects. The sizes of the sub-samples are allowed to relate to the treatment assignment, the assignment probability, the total sample size, the anticipated distributions of outcome and compliance, and the cost parameters of the study. The CSS design methods relate to prior work (i) on two-phase designs in which a covariate is sub-sampled, and (ii) on causal inference, because the sub-sampled post-randomization compliance behavior is not the true covariate of interest. For each CSS design, we develop efficient estimation of treatment efficacy under binary outcome and all-or-none observed compliance. Then, we derive a minimal cost CSS design that achieves a required precision for estimating treatment efficacy. We compare the properties of the CSS design to those of conventional protocols in a study of patient choices for medical care at the end of life.

KEY WORDS: Causal inference; Compliance sub-sampling; Double sampling; Noncompliance; Optimal design, Principal stratification.

1. Introduction

In studies with human subjects, treatment noncompliance is inevitable because subjects often do not take the assigned treatments. In analyzing data from randomized studies with noncompliance, the most common method has been intention-to-treat (ITT) analysis, that is, comparison of subjects' outcomes based on randomized treatment assignment arm alone. More recently, there has been growing literature using ideas of potential outcomes (Rubin, 1974, 1977, 1978) to estimate treatment efficacy under partial noncompliance better than ITT analyses when some reasonable assumptions hold (e.g., Sommer and Zeger, 1991; Baker and Lindeman, 1994; Imbens and Rubin, 1994; Robins and Greenland, 1994; Angrist, Imbens, and Rubin, 1996; Goetghebeur and Molenberghs, 1996; Balke and Pearl, 1997; Robins, 1998; Rubin, 1998; Frangakis and Rubin, 1999, Baker, 2000).

The recent methods that use compliance data suggest, in turn, the potential for better designs. Zelen (1979) discussed randomizing patients and then asking for their consent; Jo (1999) discussed using unbalanced probability of assignment; and Barnard et al. (2001) studied the role of prospectively recording covariates under noncompliance. However, the role of more flexible designs that plan specifically for analysis of treatment efficacy has not been explored.

We consider the framework of a simple two-arm randomized study for the comparison of a standard versus a new treatment. We assume all-or-none compliance behavior, in which each person takes either the standard or the new treatment, and the investigator can measure compliance, e.g., by analyzing blood for drug traces, or by interviewing patients. Examples are studies that randomize the encouragement to get a new screening test (e.g., for cancer) and where the treatments actually received are the taking or not taking the screen.

Studies with noncompliance are also frequent in health policy, and a recent one was conducted on Advance Directive (AD) forms (Dexter et al., 1998). AD forms 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) (e.g., Emanuel et al., 1991). AD are designed to increase patient autonomy and enjoy support by ethicists, physicians, and patients (e.g., McIntyre, 1992), but, in practice, very few eligible patients complete AD forms and very few physicians discuss the role of AD with their patients. A hypothesis has been that if physicians discuss the role of AD to their patients, this would cause substantially more patients to complete AD (Miles et al., 1996). In Sec. 5, we illustrate design and associated analyses methods to address this hypothesis using data from Dexter et al. (1998), where the randomized assignment is reminders encouraging physicians to discuss AD with their patients; actual treatment is whether or not the physician discussed AD with their patient; and the outcome is patient completion of AD.

More generally, we address studies where noncompliance to the assigned treatment is anticipated to be a serious problem, in the sense that many noncompliers can be different from compliers in terms of the outcome. Two observations are central for the further planning of such a study. First, to estimate efficacy of treatment under partial noncompliance, obtaining data on compliance behavior is necessary. Second, often there is cost in measuring compliance behavior for an individual.

We propose to use "compliance sub-sampling" designs where, after subjects are assigned treatment, compliance behavior is measured for only representative sub-samples of subjects in each assignment arm. We allow the sizes of the sub-samples to be related to the assignment arm, the assignment probabilities, the total sample size, the anticipated parameters of the distribution of outcome and compliance, and the study's cost parameters. By allowing such relations, we obtain a class of compliance sub-sampling designs that balance cost and precision to estimate treatment efficacy.

Designs that sub-sample certain variables date at least to Neyman (1938) and Cochran (1963) in the survey literature. Sub-sampling has also been discussed in epidemiologic con-

texts, for example, by Prentice (1989) for case-cohort designs, for classification and measurement error (e.g., Carroll and Wand, 1991; Baker, Connor, and Kessler, 1998, and references therein), and more recently in the context of missing covariates (e.g., Robins, Rotnitzky, and Zhao, 1994; Breslow and Chatterjee, 1999; Lawless, Kalbfleisch, and Wild, 1999). Most of the recent methods have focused on the analysis taking a design as given, not on the planning of specific sub-sampling designs. To our knowledge, there has been no work on sub-sampling of compliance data in comparative studies, specifically, on estimation and optimal planning.

In the next section, we formalize our problem in terms of its three components: (i) the epidemiologic model and definition of treatment efficacy; (ii) the study costs; and (iii) the proposed design. Section 3 discusses efficient estimation of treatment efficacy given any design in our class. In Section 4, we consider all compliance sub-sampling designs that achieve a required precision for estimating treatment efficacy and, among those, we derive a design that minimizes the study's cost. Section 5 illustrates comparisons of our design to a previous one from the AD study introduced earlier. Section 6 provides concluding remarks.

2. Problem

2.1 Epidemiologic components.

To compare a new and a standard treatment, we consider a study that can enroll consenting subjects and assign each to treatment arm $z = 0$ or 1, for standard and new treatment respectively. For the epidemiologic components of our problem, we adopt the modern framework of potential outcomes for noncompliance as introduced by Baker and Lindeman (1994), Imbens and Rubin (1994), and Angrist et al., (1996).

For subject i, let $Y_i(z)$ be the outcome when the subject is enrolled and assigned treatment arm z, for $z = 0, 1$. Let $D_i(z)$ be the treatment actually received: 0 for standard and 1 for new when the subject is assigned treatment z, for $z = 0, 1$. With respect to potential treatment received, then, subject i can be classified into one of four categories: a never-taker, denoted

by $C_i = n$, that is, a subject who, in the context of this study, would never take the new treatment whether assigned it or the standard treatment $(D_i(0) = D_i(1) = 0)$; an always-taker, denoted by $C_i = a$, that is, as subject who would always take the new treatment no matter where assigned $(D_i(0) = D_i(1) = 1)$; a complier, denoted by $C_i = c$, that is, a subject who would comply under either assignment $(D_i(0) = 0$ and $D_i(1) = 1)$; and a defier, denoted by $C_i = d$, a subject who would take the opposite treatment no matter the assignment $(D_i(0) = 1$ and $D_i(1) = 0)$.

The compliance status C_i of a subject is not always directly observed because the receipts $D_i(0)$ and $D_i(1)$ are not usually both known. We make two standard assumptions (Baker and Lindeman, 1994, Angrist et al., 1996), which facilitate inference on the compliance groups.

ASSUMPTION 1. *(Monotonicity of compliance). There are no defiers.*

ASSUMPTION 2. *(Exclusion restrictions). If subject i is an always-taker or never-taker then $Y_i(1) = Y_i(0)$ so that assignment has no effect on outcome.*

Assumption 1 is plausible rather generally because it excludes unlikely patterns of compliance behavior. Assumption 2 excludes effects of assignment on outcome for subjects who get the same treatment regardless of assignment. It is expected to hold when noncompliance occurs soon after assignment because then, within the defined groups of "always-takers" and "never-takers", different assignment results in practically the same degree of actual exposure to "new" treatment (for always-takers) and "standard" treatment (for never-takers).

The compliance status C_i is a pre-assignment characteristic whose value, by definition, does not change in this study no matter the treatment assignment. Therefore, comparisons of outcomes by assignment arm within compliance status groups are causal effects in the sense of Rubin (1978). One such comparison is the difference between the average outcome among compliers when assigned new versus standard treatment, $E(Y_i(1)|C_i = c) - E(Y_i(0)|C_i = c)$,

namely the complier average causal effect (CACE) (Baker and Lindeman, 1994; Imbens and Rubin, 1994). By Assumption 1, the compliers are the only subgroup within which outcomes in this study exist (although are not necessarily observed) under receipt of new treatment and under receipt of standard treatment. Therefore, as in recent applications (e.g., Baker, 2000; Barnard et al., 2001), we take CACE to be the causal estimand of treatment efficacy, our primary interest.

In what follows, we assume that $Y_i(1)$ and $Y_i(0)$ are binary outcomes, and denote the outcome probabilities conditionally on compliance status, $\Pr(Y_i(z) = 1|C_i = n)$, $\Pr(Y_i(z) = 1|C_i = a)$, and $\Pr(Y_i(z) = 1|C_i = c)$, by β_n, β_a, and $\beta_{c,z}$, respectively, where, by Assumption 2, the first two probabilities are not a function of assignment z. Denote the probabilities of compliance status $\Pr(C_i = n)$, $\Pr(C_i = a)$, and $\Pr(C_i = c)$, by ω_n, ω_a, and $\omega_c = 1 - \omega_a - \omega_n$ respectively, and refer to $\theta^* := (\omega_n, \omega_a, \beta_n, \beta_a, \beta_{c,1}, \beta_{c,0})$ as the canonical parameter.

Our arguments can be easily extended to allow for covariates, but to focus on main arguments, we do not address the case of covariates here. For a related point, see Section 6.

2.2 Cost components.

The study has fixed start-up costs and costs per subject (marginal). Fixed start-up costs are those that, if the study is carried out, remain roughly the same regardless of design and number of subjects enrolled, and such costs do not play a role for our purpose. For the per subject costs, it is useful to partition the expenditures in components in reverse order of follow-up as follows.

At the end of the study, a cost per subject is incurred by procedures to measure the outcome; we denote such a cost by $c^{(\text{out})}$. Also, a cost is incurred by procedures by which a subject's compliance behavior is measured. Depending on the study's context, such procedures can include, for example, laboratory testing or interviews of subjects by trained personnel. Most often, compliance behavior can be assessed reliably only if these procedures are performed during, not after, the study. We denote the per subject cost to measure compliance by $c^{(\text{cpl})}$.

The remaining cost per subject includes the cost of follow-up and the cost of providing the assigned treatment. This cost can be different in the two arms. We denote this average per subject cost by $c_0^{(\text{arm})}$ under standard and by $c_1^{(\text{arm})}$ under new treatment assignment.

2.3 Compliance sub-sampling design and goal.

To address the tradeoff between the cost incurred versus the information gained in measuring compliance to estimate treatment efficacy, we consider designs where the investigator measures compliance behavior in only a sub-sample of subjects. Because compliance is measured after treatment assignment, sub-sampling can be done separately in the two assignment arms. Such a "compliance sub-sampling" (CSS) design is demonstrated in Fig. 1. Specifically, the design first enrolls a total of n subjects. Each subject i is assigned independently to either the new $(Z_i = 1)$ or the standard $(Z_i = 0)$ treatment arm with probability $\Pr(Z_i = 1) = \lambda_1^{(\text{arm})}$; $(\lambda_0^{(\text{arm})} = 1 - \lambda_1^{(\text{arm})})$. Then, each subject i is either selected $(S_i = 1)$ by the investigator for compliance measurement procedures or not selected $(S_i = 0)$, where we assume that selection within assignment arm $z = 0, 1$ is done independently with probabilities $\lambda_z^{(\text{cpl})} = \Pr(S_i = 1 | Z_i = z)$. Outcome is measured on all subjects at the end of the study. [Figure 1 about here]

The class of CSS designs contains full compliance sampling (that is, designs that measure compliance on all subjects, as in Jo, 1999) as a special case. Importantly, CSS provides a richer choice of designs, where the sub-sampling fractions $(\lambda_0^{(\text{cpl})}, \lambda_1^{(\text{cpl})})$, as well as the total sample size n and assignment arm fractions $\lambda_1^{(\text{arm})}$ are under the investigator's control. In the next section, we show how each CSS design in our class provides data to estimate the efficacy CACE efficiently with maximum likelihood estimation (MLE). Finding optimal designs is discussed in Sec. 4.

3. Estimation given design.

The canonical parameter θ^* of the model of Sec. 2.1, in particular, CACE $(= \beta_{c,1} - \beta_{c,0})$, is not directly estimable because the compliance status C_i is missing for some subjects. It is

useful, thus, to provide links between CACE and the ITT effect of assignment on the outcomes, ITT $:=$ Pr$(Y_i(1) = 1) -$ Pr$(Y_i(0) = 1)$. Note that for any pre-assignment covariate X_i we have the decomposition ITT $= \sum_x$ ITT$_x\, t(x)$, where ITT$_x$ is the ITT effect within the subclass of subjects with covariate value x, and $t(x)$ is the proportion of subjects with covariate value equal to x. Because the compliance status C_i is a pre-assignment covariate (Sec. 2.1) we can apply the above decomposition to the compliance status, and, under Assumptions 1 and 2, which rule out ITT effects among always-takers and never-takers, we obtain ITT $=$ CACE $\cdot\ \omega_c$, giving CACE $=$ ITT$/\omega_c$ (Baker and Lindeman, 1994; Angrist et al., 1996). These arguments indicate a convenient reparameterization that facilitates finding the MLE.

Let $Y_i^{obs} = Y_i(Z_i)$, the observable outcome. In our CSS design, ITT is equal to Pr$(Y_i^{obs} = 1|Z_i = 1) -$ Pr$(Y_i^{obs} = 1|Z_i = 0)$. Also, the probability of complier ω_c is given by $\omega_c =$ Pr$(D_i^{obs} = 1|Z_i = 1) -$ Pr$(D_i^{obs} = 1|Z_i = 0)$, where $D_i^{obs} = D_i(Z_i)$, the observable treatment receipt. In our design, the receipt D_i^{obs} is not always measured either, so we will use some additional relations. By the law of total probability, we obtain that Pr$(D_i^{obs} = 1|Z_i = 1) = \sum_{y=0,1} \pi_{y,1}^{\text{(out)}} \pi_{y,1}^{\text{(cpl)}}$, where $\pi_{y,1}^{\text{(out)}} =$ Pr$(Y_i^{obs} = y|Z_i = 1)$ and $\pi_{y,z=1}^{\text{(cpl)}} =$ Pr$(D_i^{obs} = 1|Y_i^{obs} = y, Z_i = 1)$. Similarly, Pr$(D_i^{obs} = 1|Z_i = 0) = \sum_{y=0,1} \pi_{y,0}^{\text{(out)}} \pi_{y,0}^{\text{(cpl)}}$, where $\pi_{y,0}^{\text{(out)}} =$ Pr$(Y_i^{obs} = y|Z_i = 0)$ and $\pi_{y,z=0}^{\text{(cpl)}} =$ Pr$(D_i^{obs} = 1|Y_i^{obs} = y, Z_i = 0)$. Then,

$$\text{CACE} = \frac{\pi_{y=1,1}^{\text{(out)}} - \pi_{y=1,0}^{\text{(out)}}}{\sum_{y=0,1} \pi_{y,1}^{\text{(out)}} \pi_{y,1}^{\text{(cpl)}} - \sum_{y=0,1} \pi_{y,0}^{\text{(out)}} \pi_{y,0}^{\text{(cpl)}}}.$$

Under the model of Sec. 2.1, the parameter $\theta := \{\pi_{1,z}^{\text{(out)}}, \pi_{y,z}^{\text{(cpl)}}, z, y = 0, 1\}$ is in a 1-1 relation with the parameter θ^* (see Appendix) because θ and θ^* are different specifications of the joint distribution of the compliance status and potential outcomes $(C_i, Y_i(z))$, for $z = 0$ or 1, under the assumptions of Sec. 2.1 and the design. Therefore, the MLE of CACE under the model of Sec. 2.1 is CACE$(\widehat{\theta})$ where $\widehat{\theta}$ is the MLE of the parameters θ.

Let n_{yzsd} be the number of subjects with observed: assignment arm z; treatment received

d; outcome y; and compliance sub-sampling selection status $S = s$, where $s = 1$ for selection and 0 for no selection. For $z = 0, 1$ and with $+$ indicating summation over the corresponding index, we obtain the MLE $\widehat{\theta}$ of θ as: $\widehat{\pi}^{(\text{out})}_{1,z} := n_{1z++}/n_{+z++}$, $\widehat{\pi}^{(\text{cpl})}_{0,z} := n_{0z11}/n_{0z1+}$, $\widehat{\pi}^{(\text{cpl})}_{1,z} := n_{1z11}/n_{1z1+}$, and CACE$(\widehat{\theta})$ is the formula for CACE above with θ substituted by $\widehat{\theta}$, if the canonical probability estimates $\widehat{\theta}^*$ that result from inverting the above estimates using the expressions in the Appendix are all in [0,1]. If such a component of $\widehat{\theta}^*$ is not in [0,1], then its MLE is the corresponding boundary value, and $\widehat{\theta}$ is obtained by substituting the corrected $\widehat{\theta}^*$ in the relations given in the Appendix. (By Slutzky's theorem, such a correction gives the same distributional properties for CACE$(\widehat{\theta})$ as those discussed below in large enough samples, if the true parameters are in (0,1).). Denote $\partial\text{CACE}(\theta)/\partial\pi^{(\text{cpl})}_{y,z} := (-1)^z \pi^{(\text{out})}_{y,z}\,\text{ITT}\,(\omega_c)^{-2}$, and $\partial\text{CACE}(\theta)/\partial\pi^{(\text{out})}_{1,z} := (-1)^z\{(\pi^{(\text{cpl})}_{1,z} - \pi^{(\text{cpl})}_{0,z})\text{ITT} - \omega_c\}(\omega_c)^{-2}$. Then, given fixed design probabilities $\lambda := (\lambda^{(\text{arm})}_1,\ \lambda^{(\text{cpl})}_0,\ \lambda^{(\text{cpl})}_1)$, we have:

RESULT 1. *Under a CSS design,* $n^{\frac{1}{2}}\left\{\text{CACE}(\widehat{\theta}) - \text{CACE}(\theta)\right\} \to N\{0, V(\theta,\lambda)\}$,

in distribution, as $n \to \infty$, *where*

$$V(\theta,\lambda) = \sum_{z=0,1} \frac{b^{(arm)}_z(\theta)}{\lambda^{(arm)}_z} + \sum_{y,z=0,1} \frac{b^{(cpl)}_{y,z}(\theta)}{\lambda^{(arm)}_z\,\lambda^{(cpl)}_z}, \tag{3.2}$$

$$b^{(arm)}_z(\theta) := \left\{\frac{\partial\text{CACE}(\theta)}{\partial\pi^{(out)}_{1,z}}\right\}^2 \pi^{(out)}_{1,z}(1-\pi^{(out)}_{1,z}),\quad b^{(cpl)}_{y,z}(\theta) := \left\{\frac{\partial\text{CACE}(\theta)}{\partial\pi^{(cpl)}_{y,z}}\right\}^2 \frac{\pi^{(cpl)}_{y,z}(1-\pi^{(cpl)}_{y,z})}{\pi^{(out)}_{y,z}}.$$

Therefore, using a CSS design, inference for the efficacy CACE, for example confidence intervals, can be based on CACE$(\widehat{\theta})$ and the estimate of the variance (3.2) obtained by replacing the unknown parameters with their corresponding sample analogues.

4. Optimal design.

If we wish to estimate CACE with a precision no smaller than, say, $1/v_0$, then a good design would be the one that minimizes the study's total cost. Unfortunately, the true precision of CACE$(\widehat{\theta})$ in Result 1 depends on the underlying parameter. Therefore, as is common to most

study planning, we obtain the optimal design based on anticipated values θ of the parameter (as might be estimated from a pilot study). The resulting design will be nearly optimal if the true and anticipated parameters are close, but, regardless of the true value of the parameter, the estimates obtained using the design's data and the method of Result 1 will be valid.

Based on the cost components of Sec. 2.2, the cost C required to recruit n persons and follow them up according to protocol λ is

$$C(n, \lambda) = n\, Q(\lambda) \quad \text{where} \quad Q(\lambda) = c^{(\text{out})} + \sum_z \lambda_z^{(\text{arm})} (\, c_z^{(\text{arm})} + \lambda_z^{(\text{cpl})}\, c^{(\text{cpl})}). \tag{4.1}$$

When we anticipate parameter values θ, Result 1 implies that in order to estimate CACE with precision no smaller than $1/v_0$ in a CSS design with total sample size n and probabilities λ, the optimal design will satisfy $v_0 = V(\theta, \lambda)/n$. Given the required precision, we now show how to find optimal values of: the sub-sampling fractions $\lambda_1^{(\text{cpl})}, \lambda_0^{(\text{cpl})}$; the assignment probability $\lambda_1^{(\text{arm})}$; and the total sample size n, that minimize the cost $C(n, \lambda)$ in (4.1).

By rewriting $n = V(\theta, \lambda)/v_0$ and replacing this expression for n in the cost equation (4.1), as indicated by the method of Lagrange multipliers, the constraint is eliminated and the problem becomes to minimize, without constraint relation, the function $F(\theta, \lambda) := V(\theta, \lambda)\, Q(\lambda)$ over λ. To do the minimization, we have derived a Gauss-Seidel algorithm (available from the authors) which, for the anticipated parameter values θ and under weak regularity conditions, finds values of λ, say λ^{opt}, that minimize the objective function $F(\theta, \lambda)$. Finally, for our required precision $1/v_0$ in estimating CACE with CACE$(\widehat{\theta})$, we obtain the sample size, n^{opt}, of the optimal design under anticipated parameters θ by substituting θ and the optimized fractions λ^{opt} in the relation $v_0 = V(\theta, \lambda^{opt})/n^{opt}$, where $V(\theta, \lambda)$ is as in (3.2).

The same methods described above can be used to optimize any class of designs that is a special case of our CSS designs, for example with the restriction (e.g., by ethical reasons), to have balanced arms $(\lambda_1^{(\text{arm})} = 0.50)$, or with the standard full compliance sampling $(\lambda_z^{(\text{cpl})} = 1.00)$ but allowing assignment arms to be unbalanced as in Jo (1999).

5. Example on Advance Directives.

5.1 Research hypothesis and pilot study.

The designs we propose are new for noncompliance and, therefore, have not yet been used. For this reason, we return to the AD problem and the study of Dexter et al. (1998) introduced in Sec. 1. Here, our goal is to use that study's results as pilot in order to design a better study to estimate the causal effect that the physician discussion of AD with their patients has on the fraction of patients who complete AD forms. The problem of assessing structural assumptions is studied by Frangakis, Rubin, and Zhou (1998).

The study randomly divided eligible physicians into four groups: one group received reminders, by means of computer messages, to discuss instructional AD with their patients; another group received reminders to discuss proxy AD; a third group received reminders to discuss both proxy and instructional AD; the fourth group received no reminders. Here, the subset of the data we have is from the control group (receiving no reminders) and the group receiving both reminders (Table 1). Simple estimates from these data show that only 3% of patients completed AD forms when their physicians were not reminded versus a 14% completion rate among patients whose doctor was reminded, giving an estimate of 11% for the ITT effect of reminder on completion of AD. However, these estimates do not address the causal effect that physicians' discussion has on patient AD completion.

To estimate the effect of AD discussion on AD completion, we place the study in the perspective of Sec. 2, where: the assignment Z_i indicates whether or not patient i's physician received reminder to discuss AD forms with the patient; $D_i(z)$ indicates whether or not patient i would receive any discussion if that patient's physician were assigned to arm z, where $z = 1$ for reminder and $z = 0$ for no reminder; and $Y_i(z)$ indicates whether or not patient i would complete any AD forms if that patient's physician were assigned arm z. Then, the effect CACE is the causal effect that reminding a physician to discuss AD has on patient completion of AD

but only for discussion-complier patients, that is, patients whose physician: (i) would not discuss AD with the patient when not reminded in this study, and (ii) would discuss AD with the patient when reminded.

In this setting, Assumption 1 is reasonable because a physician who would discuss AD even when not reminded would invariably also discuss AD when reminded. Assumption 2 means that the computer reminder to discuss AD has no effect on the AD completion for either (i) always-discussants, that is, patients whose doctor would discuss AD with them whether reminded or not, or (ii) never-discussants, that is, patients whose doctor would not discuss AD with them whether reminded or not. These conditions are reasonable, so we also make Assumption 2 throughout. [Table 1 about here]

Table 1 provides summaries of the study's data relevant to the parameter θ discussed in Sec. 3. Under Assumptions 1 and 2, and using the relations given in the Appendix between θ and θ^* of the epidemiologic model (Sec. 2.1), we obtain the following estimates: $\widehat{\omega}_a = 0.05$, $\widehat{\omega}_n = 0.74$, $\widehat{\omega}_c = 0.21$, $\widehat{\beta}_a = 0.63$, $\widehat{\beta}_n = 0.02$, $\widehat{\beta}_{c,z=0} = 0.00$, $\widehat{\beta}_{c,z=1} = 0.48$, and $\widehat{\text{CACE}}(= \widehat{\beta}_{c,z=1} - \widehat{\beta}_{c,z=0}) = 0.48$. Therefore, there are approximately 21% discussion-compliers, and the efficacy CACE is approximately 48%, which is substantially higher than the 11% ITT estimate. However, the present study had not been designed to estimate CACE, and aspects specific to treatment efficacy estimation had not been used. For example, one drawback of the design in Dexter et al. (1998) was the use of cluster randomization: when a physician was reminded, the reminder was to discuss AD to all their patients (indiscriminately); this clustered assignment could have been avoided (see next section). Also, importantly, the study did not consider possible CSS designs. Due to these and related complications discussed in Frangakis, Rubin, and Zhou (1998), we regard the current estimates as only pilot information which we use to design a new study to estimate CACE well with controlled cost. Next, we describe this projected new study, where we compare CSS to full compliance sampling designs.

5.2 Projected study comparisons.

Epidemiologic component.

Based on the pilot estimates reported above, we consider 9 conditions for the true canonical parameters of the projected study. For simplicity, across conditions we fix at the pilot estimates the baseline probabilities $\beta_n = 0.02$, $\beta_a = 0.63$, and the probability of always-discussants, $\omega_a = 0.05$; we set $\beta_{c,z=0}$ to 0.01, instead of the pilot estimate value 0.00, in order to allow for the non-degenerate case of small but positive true probability of AD completion in the control arm. We vary the other parameters around their pilot estimates: we set the proportion of discussion compliers to $\omega_c = 0.11$, 0.21 or 0.31 (giving $\omega_n = 0.84$, 0.74, or 0.64 respectively); we set CACE to $0.38, 0.48$, or 0.58 (giving $\beta_{c,1} = 0.39$, 0.49, or 0.59 respectively).

Cost component.

Specific cost information for the study of Dexter et al. (1998) was not available, but, because only relative costs are needed, we used plausible estimates based on the description of the pilot study. In the pilot study, a patient was recorded as having completed an AD form when the form was received by the hospital administration. Therefore, the cost of measuring the outcome for each patient was the time taken by the staff assistant to retrieve and record the data for each patient, a relatively fast task. Measuring compliance however, that is, finding out whether a patient "received" a discussion on AD by the physician, required individual patient confidential interview (Dexter et al., 1998). We therefore estimated that the cost $c^{(\text{cpl})}$ in such a study should be at least four times $c^{(\text{out})}$, so we set $c^{(\text{out})} = 1, c^{(\text{cpl})} = 4$. We estimated the per subject cost of assigning arm, $c^{(\text{arm})}$, as practically 0 because assignment to reminder (or not) of physicians was done by a computer program, the development and operation of which is a fixed start-up cost.

Design component and results.

For the projected study, we consider the class of CSS designs of Sec. 2.3, adapted to the

setting of the reminder study of Dexter et al. (1998) for two assignment arms – reminders or no reminders. In order for our causal model and notation $D_i(z)$ to be sufficient for this setting, we need to assume that a physician's decision to discuss or not AD with a given patient i is a function of whether or not that physician receives reminders about patient i but not about a different patient (no interference, Rubin, 1978). This assumption is plausible when, as in the original study, the "reminder" comprises repeated messages of encouragement for AD discussion for that patient, in which case a physician's non/discussion of AD with that patient when reminded is less due to memory and more due to that physician's conscious decision, based on the received encouragement, to discuss (or not) AD with that patient. To further increase precision for estimating CACE, we assume that the projected study avoids the clustered randomization of Dexter et al. (1998) by randomizing the patients, instead of the physicians, to two groups. For the patients in the first group, their physician would receive reminders to discuss AD to these patients, whereas no action would be taken regarding the patients in the second group. Under these conditions, the above randomized assignment is ignorable (Rubin, 1978) without conditioning on the physician indicators (even if different physicians have different degrees of preference in discussing AD) and the estimates and variance calculation of Sec. 3 can be used on data of the projected study without adjustments for clustering. If, however, one wishes to project the new study to use, as the original one did, randomization of patients at the cluster (physician) level instead of the patient level, then it would be necessary, with or without interference between units, to inflate the variance formulas of Sec. 3 to account for the correlation, using, for example, sampling-based (e.g., jackknife) or more model-based (Frangakis, Rubin, and Zhou, 1998) approaches.

For each of the 9 conditions for the epidemiologic parameters (the rows of Table 2), we used the methods of Sec. 4 to find the CSS design that would require the lowest cost in order to estimate CACE with a fixed precision, say v_0^{-1}. The actual value v_0 does not matter

for the relative comparisons in Table 2. Optimal sub-sampling fractions and cost of this design are given in columns 9-12 of Table 2. For the same precision in estimating CACE, we also calculated: (i) the minimum-cost CSS design among those with balanced assignment arms ($\lambda_1^{(\text{arm})} = 0.50$) (columns 5-7); (ii) the minimum-cost design among those that measure compliance on all subjects (full sampling design: ($\lambda_0^{(\text{cpl})} = \lambda_1^{(\text{cpl})} = 1.00; \lambda_1^{(\text{arm})}$ free), columns 3-4); and (iii) the minimum-cost design among designs with full compliance sampling and balanced arms ($\lambda_0^{(\text{cpl})} = \lambda_1^{(\text{cpl})} = 1.00; \lambda_1^{(\text{arm})} = 0.50$) (column 2). Sample size is free to be optimized in all designs. To reflect relative costs, all cost figures in Table 2 are in % of the cost of the optimal CSS design in column 12 when the epidemiologic parameters are equal to the estimates from the pilot study (CACE=48%, $\omega_c = 21\%$). Columns 8 and 13 are the cost of the optimal full sampling design (column 4) divided, respectively, by the costs of the two CSS designs (columns 7 and 12). [Table 2 about here]

The results of Table 2 can be summarized in three main points. First, although the assignment probabilities between the optimal CSS and optimal full-sampling designs (columns 9 and 3 respectively) do not differ much, all the optimal CSS probabilities $\lambda_1^{(\text{cpl})}$ and $\lambda_0^{(\text{cpl})}$ here are less than 50%. As a result, the optimal total sample size n^{opt} for the CSS designs, obtained by $n^{opt} = V(\theta, \lambda^{opt})/v_0$ (Sec. 4), is larger (approximately 60%) than that in the full compliance sampling designs. Thus, by optimality construction for a required precision, the extra cost of additional initial recruitment here is substantially more than offset by not measuring compliance on more than half of the subjects using CSS. Second, although there are conditions where the gain in total cost of the CSS design can be modest (e.g., column 13=120%), there are also conditions where this design gives substantial gains over the full sampling design (e.g., column 13=134%, 156%) . Third, the efficiency of the CSS design with restricted balanced assignment arms in these cases is close to that of the optimal unrestricted CSS design.

5.3 Sensitivity to parameter values.

We investigated sensitivity by considering situations where the true parameter value, say θ_0, differs from the anticipated value, say θ_a, used in calculating the designs. To do this, we fixed the anticipated parameter value θ_a to the pilot AD estimates (i.e., condition of Table 2 with CACE = 48% and Pr(compliers)=21%), and we varied the true value θ_0 as in the 9 conditions of Table 2. The new setting is depicted in Table 3.

The designs of Table 3 are the designs of Table 2 for the condition CACE=48% and Pr(compliers)=21%, and are all calculated to have a common precision v^{-1} for estimating CACE. Now, however, the true precision differs across designs because the anticipated parameter θ_a generally differs from the true value θ_0. Therefore, a better measure to compare across designs in Table 3 is the cost per unit of precision, which, by Result 1 and (4.1), is $C(n,\lambda)/\{n/V(\theta_0,\lambda)\} = F(\theta_0,\lambda)$ of Sec. 4. For each value of the true parameter θ_0 and for each design, Table 3 reports the cost per precision $F(\theta_0,\lambda)$ divided by the cost per precision of the CSS design with all probabilities λ free and when the true parameter values are CACE=48% and Pr(compliers)=21%. [Table 3 about here]

The results of Table 3 are similar to those of Table 2 (the 2nd columns of the tables, as expected, are identical). Therefore, if the true parameter differs from the anticipated parameter under which the designs are calculated, then the CSS design here is still better than the full compliance sampling design, although the gain is mostly a function of the true parameter value and to a smaller degree a function of the anticipated value.

6. Remarks.

The proposed design is applicable in studies where noncompliance is anticipated, Assumptions 1 and 2 approximately hold, and there is a cost to measuring compliance data. The implications of Assumptions 1 and 2 should be considered for each study. For example, as noted by a reviewer, in studies where subjects in the "standard arm" have no access to the new, e.g.,

experimental, treatment, there will be neither defiers nor always-takers, by design.

We presented our CSS design for a dichotomized compliance measure and outcome. Using a full compliance sampling design, Frangakis and Rubin (1999, Sec. 5) and Baker (2000), and Baker and Lindeman (2001, Sec. 7) have presented analyses methods for survival and continuous censored outcomes, and multilevel compliance, respectively, so it would be interesting to combine such analyses with our CSS design methods, in analogous ways to those presented in this paper. In a related issue mentioned by a reviewer, the goal of high precision can be on a combination of estimands, e.g. of CACE with ITT. Such goals would indicate a new objective function (or constraint), and thus a new optimal CSS design.

With noncompliance, covariates are useful in the analysis, because they can increase precision in predicting the partially unobserved compliance groups, especially when assumptions such as the exclusion restriction are relaxed (e.g., Frangakis et al., 1998, Sec. 3.2). Using covariates at the design stage also requires anticipated values of the parameters for the interactions of such covariates with the compliance status, outcomes, and assignment. When such knowledge is available, we recommend incorporating the covariates in the CSS design, to even better balance the gains in precision with potential costs of measurement.

The approach we used to adjust for noncompliance is a special case of "principal stratification" (Frangakis and Rubin, 2000) to adjust for general post-treatment variables (e.g., putative surrogate endpoints), where treatments are compared within strata of the potential, not just the observed, values of the post-treatment variables. It would, therefore, be useful to apply our design methods in studies with other post-treatment variables. Finally, there are often also incomplete outcomes due to dropout, with considerable cost in finding (double sampling) dropout subjects to measure it. Baker, Wax, and Paterson (1993) and Frangakis and Rubin (2001) have proposed methods for double sampling "time-to-event" outcomes, which would be useful to combine with CSS designs.

ACKNOWLEDGEMENT

We thank the Editor, the Associate Editor, anonynous reviewers, Ron Brookmeyer, Booil Jo, and Scott Zeger for constructive comments, and the National Institute of Child Health and Human Development (R01 HD38209), the National Institute for Drug Abuse (R01 DA10184), and the H.-C. Yang Memorial Faculty Fund for partial financial support.

REFERENCES

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with Discussion). *Journal of the American Statistical Association* **91**, 444–472.

Baker, S. G. (2000). Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. *Journal of the American Statistical Association* **95**, 43–50.

Baker, S. G., Connor, R. J., and Kessler, L. G. (1998). The partial testing design: a less costly way to test equivalence for sensitivity and specificity. *Statistics in Medicine* **17**, 2219-2232.

Baker, S. G. and Lindeman, K. S. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* **13**, 2269–2278.

Baker, S. G. and Lindeman, K. S. (2001). Rethinking historical controls. Forthcoming in *Biostatistics*.

Baker, S. G., Wax, Y., and Patterson, B. H. (1993). Regression analysis of grouped survival data: informative censoring and double sampling. *Biometrics*, **49**, 379–389.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 93, 929–934.

Barnard, J., Frangakis, C. E., Hill, L., and Rubin, D. B. (2001). School Choice in NY city: a Bayesian analysis of an imperfect randomized experiment. *Case Studies in Bayesian Statistics*, (with discussion), C. Gatsonis, et al. (eds.). New York: Springer-Verlag, in press.

Breslow, N. and Chatterjee, N. (1999). Design and analysis of two phase studies with binary outcome applied to wilms tumor prognosis. *Applied Statistics*, in press.

Carroll, R. and Wand, M. (1991). Semiparametric estimation in logistic regression measurement error models. *Journal of the Royal Statistical Society B* **53**, 573–585.

Cochran, W. (1963). Sampling techniques. New York: Wiley.

Dexter, P., et al. (1998). Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directives. *Annals of Internal Medicine* **128**, 102–110.

Emanuel, L. L., et al. (1991). Advance Directives for medical care – a case for greater use. *New England Journal of Medicine* **324**, 889–895.

Frangakis, CE, and Rubin, DB (2001). A note on addressing the idiosyncrasy in estimating survival curves using double-sampling in the presence of self-selected right censoring. *Biometrics*, in press.

Frangakis, CE, and Rubin, DB (2000). The defining role of "principal stratification" in causal inference: adjusting for post-treatment variables. *Proceedings of the Epidemiology Section*, American Statistical Association (invited paper) 2000.

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

Frangakis, C. E., Rubin, D. B., and Zhou, X.-H. (1998). The clustered-encouragement-design. In *Proceedings of the Biometrics Section, American Statistical Association*, pp. 71–79.

Goetghebeur, E. and Molenberghs, G. (1996). Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association* **91**, 928–934.

Imbens, G. W. and Rubin, D. B. (1994). Causal inference with instrumental variables. Discussion paper # 1676. Cambridge, MA: Harvard Institute of Economic Research.

Jo, B. (1999). Power to detect intervention effects in randomized trials with noncompliance. Technical Report, Graduate school of education and information studies, UCLA.

Lawless, J. Kalbfleisch, J., and Wild, C. (1999) Semiparametric methods for response selective and missing data problems in regression. *Journal of the Royal Statistical Society B* **61**, 413–438.

Miles, S., Koepp, P. R, and Weber E. P. (1996). Advance end of life treatment planning. A research review. *Archives of Internal Medicine* **156**, 1062–68.

McIntyre, K. M. (1992). Implementation of advance directives. For physicians, a legal dilemma becomes

an ethical imperative. *Archives of Internal Medicine* **152**, 925–929.

Neyman, J. (1938). Contribution to the theory of sampling from human populations. *Journal of the American Statistical Association* **33**, 101–116.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine* **17**, 269–302.

Robins, J. M. and Greenland, S. (1994). Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* **89**, 737–749.

Robins, J. M., Rotnitzky, A., and Zhao, L.-P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

Rubin, D. B. (1977). Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics* **2**, 1-26.

Rubin, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6**, 34–58.

Rubin, D. B. (1998). More powerful randomization-based p-values in double-blind trials with noncompliance. *Statistics in Medicine* **17**, 371–387.

Sommer, A. and Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.

Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine* **300**, 1242–1245.

APPENDIX A: RELATION BETWEEN PARAMETERIZATIONS OF SEC. 2 AND SEC. 3.1.

We relate the parameters of Sec. 3.1 to the canonical parameters of Sec. 2. For the probabilities of observable outcome, we have $\Pr(Y_i^{obs} = 1|Z_i) = E\{\Pr(Y_i^{obs} = 1|Z_i, C_i)|Z_i\}$, so $\pi_{1,0}^{(\text{out})} =$

$\omega_a\beta_a + \omega_n\beta_n + \omega_c\beta_{c,0}$ and $\pi^{(\text{out})}_{1,1} = \omega_a\beta_a + \omega_n\beta_n + \omega_c\beta_{c,1}$. For the conditional probabilities of observable treatment receipt, we have $\Pr(D_i^{obs} = 1|Y_i^{obs} = y, Z_i) = E\{\Pr(D_i^{obs} = 1|Y_i^{obs} = y, Z_i, C_i)|Y^{obs} = y, Z_i\}$. Therefore,

$$\pi^{(\text{cpl})}_{y,1} = \frac{\omega_a(\beta_a)^y(1-\beta_a)^{1-y} + \omega_c(\beta_{c,1})^y(1-\beta_{c,1})^{1-y}}{(\pi^{(\text{out})}_{1,1})^y(1-\pi^{(\text{out})}_{1,1})^{1-y}},\ \pi^{(\text{cpl})}_{y,0} = \frac{\omega_a(\beta_a)^y(1-\beta_a)^{1-y}}{(\pi^{(\text{out})}_{1,0})^y(1-\pi^{(\text{out})}_{1,0})^{1-y}},$$

for $y = 0, 1$. The canonical parameters as functions of the reparameterization in Sec. 3.1 are obtained, by inversion, as $\omega_n = 1 - \sum_{y=0,1} \pi^{(\text{out})}_{y,1}\pi^{(\text{cpl})}_{y,1}$; $\omega_a = \sum_{y=0,1} \pi^{(\text{out})}_{y,0}\pi^{(\text{cpl})}_{y,0}$; $\omega_c = 1 - \omega_n - \omega_a$; $\beta_n = (1 - \pi^{(\text{cpl})}_{1,1})\pi^{(\text{out})}_{1,1}\omega_n^{-1}$; $\beta_a = \pi^{(\text{cpl})}_{1,0}\pi^{(\text{out})}_{1,0}\omega_a^{-1}$; $\beta_{c,1} = \{\pi^{(\text{out})}_{1,1}\pi^{(\text{cpl})}_{1,1} - \pi^{(\text{cpl})}_{1,0}\pi^{(\text{out})}_{1,0}\}\omega_c^{-1}$; and $\beta_{c,0} = \{(1 - \pi^{(\text{out})}_{1,0})\pi^{(\text{cpl})}_{1,0} - \pi^{(\text{out})}_{1,1}(1 - \pi^{(\text{cpl})}_{1,1})\}\omega_c^{-1}$.

Figure 1. Compliance sub-sampling design. Solid boxes represent observed information; dashed boxes represent unobserved information.

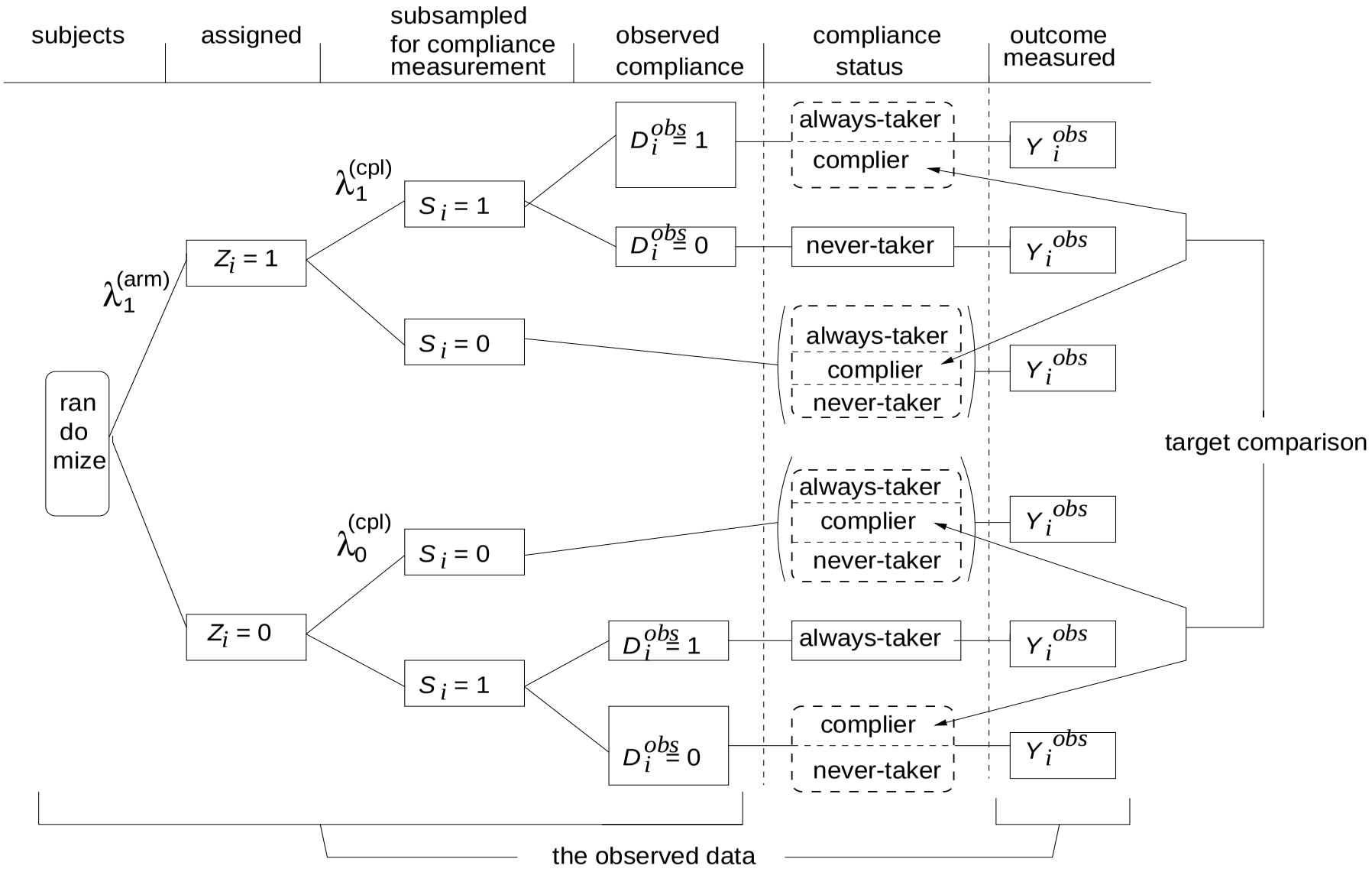
subjects
assigned
subsampled for compliance measurement
observed compliance
compliance status
outcome measured
randomize
$Z_i = 1$
$Z_i = 0$
$\lambda_1^{(arm)}$
$\lambda_1^{(cpl)}$
$\lambda_0^{(cpl)}$
$S_i = 1$
$S_i = 0$
$S_i = 0$
$S_i = 1$
$D_i^{obs} = 1$
$D_i^{obs} = 0$
$D_i^{obs} = 1$
$D_i^{obs} = 0$
always-taker
complier
never-taker
always-taker
complier
never-taker
always-taker
complier
never-taker
always-taker
complier
never-taker
Y_i^{obs}
Y_i^{obs}
Y_i^{obs}
Y_i^{obs}
Y_i^{obs}
Y_i^{obs}
target comparison
the observed data

Table 1: Patient characteristics in our sample from the study of Advance Directives. Estimates are based on sampling-theory ratio estimation for cluster sampling (Cochran, 1963, p. 30). (To generalize to the larger population of comparable patients, the finite population correction is set here to 1.)

	Control assignment		Reminder assignment		Difference between assignments		
Doctors (no.)	26		25				
Patients (no.)	158		175				
Completed AD, %	3.2	(1.3)	14.3	(4.4)	11.1	(4.6)	[2.4]
Had discussed AD, %							
among completers	100.0	−−	92.0	(4.7)			
among non-completers	2.0	(1.0)	14.7	(3.4)			
among all	5.1	(1.7)	25.7	(5.5)	21.6	(5.8)	[3.7]

Estimates are percents (se). Numbers in brackets are ratios of estimated percents over their standard errors.

True parameter value[1]	Full Sampling			Compliance Sub-sampling								
	$(\lambda_1^{(arm)} = 0.50)$	$(\lambda_1^{(arm)}$ free)		$(\lambda_1^{(arm)} = 0.50, \lambda_0^{(cpl)}, \lambda_1^{(cpl)}$ free)				$(\lambda_1^{(arm)}, \lambda_0^{(cpl)}, \lambda_1^{(cpl)}$ free)				
CACE 58%	Cost_1	$\lambda_1^{(arm)}$	Cost_2	$\lambda_0^{(cpl)}$	$\lambda_1^{(cpl)}$	Cost_3	$100\frac{\text{Cost}_2}{\text{Cost}_3}$	$\lambda_1^{(arm)}$	$\lambda_0^{(cpl)}$	$\lambda_1^{(cpl)}$	Cost_4	$100\frac{\text{Cost}_2}{\text{Cost}_4}$
(col.#): (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Pr(complier) 11%	401.0	0.58	391.5	0.32	0.50	327.1	119.7	0.56	0.36	0.45	325.6	120.2
21%	137.9	0.62	130.1	0.29	0.54	111.0	117.2	0.60	0.36	0.46	109.4	118.9
31%	76.1	0.66	69.3	0.26	0.55	59.8	115.9	0.64	0.37	0.44	58.1	119.3
CACE 48%												
Pr(complier) 11%	412.6	0.58	402.7	0.24	0.40	297.0	135.6	0.56	0.27	0.36	295.2	136.4
(pilot row) 21%	142.1	0.62	134.0	0.21	0.43	101.8	131.6	0.60	0.27	0.37	100.0	134.0
31%	78.5	0.66	71.5	0.19	0.44	55.1	129.8	0.64	0.28	0.36	53.3	134.1
CACE 38%												
Pr(complier) 11%	422.7	0.57	414.0	0.17	0.30	259.5	159.5	0.56	0.20	0.28	257.7	160.7
21%	143.4	0.62	136.1	0.16	0.34	88.7	153.4	0.60	0.20	0.29	87.0	156.4
31%	78.4	0.65	72.1	0.14	0.35	47.8	150.8	0.63	0.20	0.28	46.2	156.1

Notes:

1. Anticipated and true epidemiologic parameter values are equal within each row.

2. All cost figures are relative to 100 units of cost for the optimal CSS design under the condition with parameters equal to the pilot estimates from the AD study: CACE=48% and $\Pr(\text{complier}) = 21\%$.

3. Each design shown is optimized over any free design parameters in its class (see also Secs. 4 and 5.2)

Table 2: Designs for efficacy estimation. Projected relative costs.

True parameter value[1]	Full Sampling		Compliance Sub-sampling			
	$(\lambda_1^{(arm)} = 0.50)$	$(\lambda_1^{(arm)}$ free)	$(\lambda_1^{(arm)} = 0.50, \lambda_0^{(cpl)}, \lambda_1^{(cpl)}$ free)		$(\lambda_1^{(arm)}, \lambda_0^{(cpl)}, \lambda_1^{(cpl)}$ free)	
CACE 58%	Cost_1^*	Cost_2^*	Cost_3^*	$100\frac{\text{Cost}_2^*}{\text{Cost}_3^*}$	Cost_4^*	$100\frac{\text{Cost}_2^*}{\text{Cost}_3^*}$
(col.#): (1)	(2)	(3)	(4)	(5)	(6)	(7)
Pr(complier) 11%	401.0	395.0	334.8	118.0	333.8	118.3
21%	137.9	130.1	112.7	115.4	111.0	117.2
31%	76.1	69.6	60.5	115.0	59.0	118.0
CACE 48%						
Pr(complier) 11%	412.6	406.2	298.5	136.1	297.8	136.4
(pilot row) 21%	142.1	134.0	101.8	131.6	100.0	134.0
31%	78.5	71.8	55.2	130.0	53.5	134.2
CACE 38%						
Pr(complier) 11%	422.7	418.6	265.9	157.4	266.1	157.3
21%	143.4	136.2	90.4	150.7	88.7	153.6
31%	78.4	72.3	48.9	147.9	47.2	153.2

Notes:

1. Anticipated epidemiologic parameter value is fixed at the pilot estimates, CACE=48% and $pr(\text{complier}) = 21\%$.

2. Cost* is the cost per unit of precision ($F(\theta_0, \lambda)$, see Sec. 5.3) relative to 100 units of the cost per precision for the optimal CSS design under the setting with CACE=48% and $\Pr(\text{complier}) = 21\%$.

Table 3: Sensitivity to deviations of true from anticipated parameters.