Biostatistics (2008), **9**, 2, *pp.* 277–289
doi:10.1093/biostatistics/kxm027
Advance Access publication on August 6, 2007

Principal stratification with predictors of compliance for randomized trials with 2 active treatments

JASON ROY*

Center for Health Research, Geisinger Health System, Danville, PA, USA
jaroy@geisinger.edu

JOSEPH W. HOGAN

Center for Statistical Sciences, Brown University, Providence, RI, USA

BESS H. MARCUS

Centers for Behavioral and Preventive Medicine and Division of Cardiology, Brown Medical School and The Miriam Hospital, Brown University, Providence, RI, USA

SUMMARY

In behavioral medicine trials, such as smoking cessation trials, 2 or more active treatments are often compared. Noncompliance by some subjects with their assigned treatment poses a challenge to the data analyst. The principal stratification framework permits inference about causal effects among subpopulations characterized by potential compliance. However, in the absence of prior information, there are 2 significant limitations: (1) the causal effects cannot be point identified for some strata and (2) individuals in the subpopulations (strata) cannot be identified. We propose to use additional information—compliance-predictive covariates—to help identify the causal effects and to help describe characteristics of the subpopulations. The probability of membership in each principal stratum is modeled as a function of these covariates. The model is constructed using marginal compliance models (which are identified) and a sensitivity parameter that captures the association between the 2 marginal distributions. We illustrate our methods in both a simulation study and an analysis of data from a smoking cessation trial.

Keywords: Bounds; Causal effect; Latent class model; Mediation; Noncompliance; Potential outcomes.

1. INTRODUCTION

In behavioral medicine trials, new interventions are usually tested relative to a standard or existing behavioral or pharmacologic therapy. A specific example is the commit to quit (CTQ) trials (Marcus *and others*, 1999, 2005), comprising 2 longitudinal follow-up studies of supervised exercise to promote smoking cessation. Each study had 2 treatment arms. All participants received cognitive-behavioral smoking

*To whom correspondence should be addressed.

cessation therapy. For those in the intervention arm, cognitive-behavioral therapy (CBT) was augmented by an individualized, supervised exercise program. In order to equalize contact hours between the 2 arms, CBT for those in the control arm was augmented by a wellness education program that included lectures, films, handouts, and discussions covering issues such as healthy eating and prevention of cardiovascular disease. Hence, the comparison is between standard therapy augmented by wellness education and standard therapy augmented by an exercise regimen.

A key complication in drawing inference about causal effects in any trial is that compliance is rarely perfect. A standard approach is to estimate the intention-to-treat (ITT) effect. The ITT effect describes the benefit of being randomized to treatment. ITT can be used, for example, to determine the impact that recommending (or encouraging; Hirano *and others*, 2000) a particular treatment over an alternative treatment would have, on average, in the population. However, one treatment might appear superior because it is better tolerated. For example, one could imagine a treatment that fewer people would comply with but that has superior results among those that do comply. If this information was known, the recommended treatment could potentially be tailored to individual subjects. Thus, in addition to knowing the ITT effect, it would be useful to know which treatment would result in better outcomes among subjects who would comply with and receive either intervention and which characteristics are predictive of compliance.

In trials of an active treatment versus placebo (or no treatment), it is possible to recover causal effects under some reasonably mild assumptions. For all-or-none compliance situations, the method of instrumental variables can be particularly useful (Angrist *and others*, 1996). Causal effects among compliers (subjects who would take treatment if offered) are identifiable under the assumption that there are no subjects who would take the active treatment if randomized to the control arm but not to the treatment arm (no defiers). This assumption is reasonable in trials where the control group does not have access to the active treatment. In such settings, the instrumental variables estimator is equivalent to the estimator from certain structural mean models (Robins, 1994; Goetghebeur and Lapp, 1997; Robins and Rotznitzky, 2004). Structural mean models can also be used when compliance is continuous and if there are interactions between the causal effect and the baseline covariates. However, if it is unreasonable to assume there are no defiers, the causal parameters are generally not identifiable without structural assumptions; several authors have derived bounds on the causal effects (Robins, 1989; Manski, 1990; Balke and Pearl, 1997; Joffe, 2001).

For comparisons of 2 active treatments, the principal stratification framework (Frangakis and Rubin, 2002) can be used to define and infer causal parameters of interest, such as the effect of the exercise intervention compared with standard therapy among subjects who would comply with either intervention. Direct implementation of the principal stratification approach has several limitations. First, causal effects of interest would not be point identified (Cheng and Small, 2006). Further, it would not provide information about characteristics of subjects in each subpopulation (stratum). To address these issues, we identify causal parameters of interest, up to a sensitivity parameter, through the use of baseline covariates that are predictive of compliance. Including covariates in the model is not straightforward as care has to be taken to ensure the marginal compliance distributions are compatible with the joint distribution. In addition, our model clearly separates parameters that can be identified from the data from those that cannot. Finally, we illustrate how a constraint on the joint distribution of the potential compliance variables can be integrated into the methodology. This approach provides investigators with additional useful information, beyond just the ITT effect.

The remainder of the paper is organized as follows: In Section 2, we describe the smoking cessation trial. We introduce the notations and assumptions in Section 3. In Section 4, we present the compliance models. The likelihood, prior, and posterior distributions are described in Section 5. Data from the CTQ smoking cessation trial are analyzed in Section 6. There are concluding remarks in Section 7. A simulation study, derivation of the observed-data likelihood, and some additional results are presented in the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org).

2. Data from "CTQ" trial

The first CTQ study (Marcus *and others*, 1999) was a randomized controlled trial designed to assess the efficacy of supervised vigorous exercise as an adjuvant treatment to CBT for promotion of smoking cessation among women. A primary motivation behind studying women exclusively was that in the mid-1990s, smoking prevalence rates among women were declining at a slower rate than among men (Escobedo and Peddicord, 1996). In addition, relapse rates among women were higher. The study enrolled and assigned 134 women to receive CBT plus vigorous exercise (the new treatment) and 147 to receive CBT plus a wellness education program (the control treatment). CBT represents the standard of care for smoking cessation; the wellness education was added to the control arm to equalize staff contact time between the 2 arms and to eliminate the possibility that treatment effects associated with exercising might be confounded by added staff time. The CBT program was administered to all women in group format weekly over the course of 12 weeks. The exercise program was individually tailored to each woman based on achieving a target heart rate; women exercised 3 times weekly, supervised by an exercise specialist. Women in the control arm participated in a program of supervised lectures, films, and discussions 3 times weekly and were instructed not to adopt a program of regular exercise until the study was completed. None of the women in the control arm had access to the supervised exercise program, and none in the exercise group had access to wellness classes.

The target quit date was week 5 following randomization. Cessation status was evaluated weekly over the course of 12 weeks. Cessation was assessed by self-report and, for the first week in which abstinence was reported, verified by carbon monoxide ($<$8 ppm) and saliva cotinine ($<$10 ng/mL). The primary outcome of the study was continuous abstinence during the 8 weeks after the quit date. By definition, an individual who was not present for scheduled testing at one or more occasions could not be counted as continuously abstinent.

Subjects were expected to attend 3 wellness or exercise classes per week. CBT was administered during the first exercise or wellness visit that was attended in a given week. We defined compliance based on the number of sessions actually attended during the 4 weeks prior to the quit date. Specifically, a subject was defined as compliant with their assigned treatment if she attended at least two-thirds of the classes every week and attended all 3 classes at least twice, during weeks 1–4. Compliance status is observed for all participants.

Compliance was similar in the wellness and exercise arms (41% and 43%, respectively). Continuous abstinence rates were 10.8% in the wellness arm and 19.4% in the exercise arm. These are estimates of the expected abstinence rates under the observed compliance pattern (ITT). Our goal is to estimate the treatment effect among those who would comply with either intervention.

3. Principal stratification with 2 active treatments

3.1 *Notations and assumptions*

We consider experimental trials with 2 arms—each with an active treatment. Our approach can easily be extended to 3-arm trials with 2 active treatments (where one arm is a control group), such as considered by Cheng and Small (2006). Let $Z \in \{0, 1\}$ denote a randomization indicator, where $Z = 1$ indicates randomization to the new treatment (e.g. supervised exercise plus CBT) and $Z = 0$ indicates randomization to standard therapy (e.g. wellness sessions plus CBT). Let $A_z \in \{0, 1\}$ denote compliance with *assigned* treatment under assignment z. Similarly, define $Y_z \in \{0, 1\}$ to be the outcome under assignment z. Each person has 2 potential compliance levels, A_0 and A_1, that characterize compliance under either treatment assignment; however, only $A = ZA_1 + (1 - Z)A_0$ is observed. Similarly, each subject has 2 potential outcomes, Y_0 and Y_1, with $Y = ZY_1 + (1 - Z)Y_0$ observed.

We make 3 standard assumptions for the development of analytic methods: (1) the stable unit-treatment value assumption (SUTVA), (2) randomization ($Z \perp\!\!\!\perp \{Y_0, Y_1, A_0, A_1\}$), and (3) the exclusion restriction. These assumptions have been described in detail elsewhere (Angrist *and others*, 1996). We make 2 additional assumptions.

Assumption 4. Treatment access restriction. Subjects in group $Z = z$ do not have access to the treatment assigned in arm $Z = 1 - z$, for $z = 0, 1$. This assumption holds in CTQ because subjects in the one arm were not allowed to attend the classes in the other.

Assumption 5. Monotonicity. The probability of compliance with treatment assigned by $Z = 1$ is higher among those who would comply with treatment assigned by $Z = 0$, compared to those who would not; that is $P(A_1 = 1|A_0 = 1) \geqslant P(A_1 = 1|A_0 = 0)$.

The purpose of Assumption 5 is to tighten the bounds on the causal effects because in many practical settings, they are too wide to be useful. The assumption is plausible in studies where both treatments require a similar commitment of effort and where the two treatments are not polarizing in terms of preferences. It should be noted that the assumption refers to average compliance rates and does allow for the possibility that some subjects would comply with one intervention but not with the other. Applied to CTQ, it states that compliance with exercise would be more prevalent among those who would comply with the control condition (i.e. positive correlation between A_0 and A_1). In CTQ, both interventions required traveling 3 days per week to attend the sessions. The study population only included subjects who were not exercising regularly. The time commitment for both interventions was the same. We therefore believe the assumption is plausible. However, our methods can easily be amended to exclude Assumption 5.

3.2 *Defining causal effects*

We assume that each individual belongs to one of 4 basic principal strata defined by unique combinations of (A_0, A_1). The strata comprise the set $\mathcal{S} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. Hence, the subpopulation with $S = (1, 1)$ includes only women who would comply with whichever treatment was offered; $S = (0, 1)$ denotes the population that would comply with exercise but not with the control treatment, and so forth.

In the most general terms, we are seeking the joint distributions $[(Y_0, Y_1) \mid S = s]$, for $s \in \mathcal{S}$, which characterize the causal effect of randomization to treatment within principal stratum s. Let $\pi_z^{(s)} = P(Y_z = 1|S = s)$. In our analysis of the CTQ trial, the risk difference $\pi_1^{(1,1)} - \pi_0^{(1,1)}$ is a target for inference, because it is the contrast that captures the effect of actually receiving exercise, and is therefore of direct interest from the point of view of intervention design. In addition, we are interested in the distribution of covariates, given the compliance class.

3.3 *Bounds on causal effects in the absence of covariates*

In trials with 2 active treatments, the causal effects within each basic principal stratum are not point identified from the outcome and compliance data alone under Assumptions 1–4 (Cheng and Small 2006). That is, even if the entire population was observed, the (frequentist) point estimate would only be known to lie within an interval. We derive these bounds in Section A of the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org).

Our strategy is to further reduce uncertainty about the treatment effects by using information from baseline covariates that are predictive of compliance. For example, suppose women with a low body mass index (BMI) were more likely to comply with the wellness intervention than women with a high BMI. It is realistic to believe that a woman with low BMI in arm $Z = 1$ who complied with exercise is more likely to have $S = (1, 1)$ than a woman with high BMI who complied with exercise. We formalize this in a model for (A_0, A_1) and use it to identify causal effects.

4. Utilization of compliance-predictive covariates

Assume that for each subject, we have several baseline covariates X that are associated with compliance in one or both arms. We are able (using randomization of Z) to identify the marginals $[A_0|X]$ and $[A_1|X]$, but we cannot identify the conditional association between A_0 and A_1 given X. Our approach is to specify separately the marginal distributions $[A_0|X]$ and $[A_1|X]$, use a single non-identifiable parameter ϕ to capture the association structure, and thereby characterize the joint distribution $[A_0, A_1|X]$ (see also Heagerty, 2002; Dominici *and others*, 2006).

4.1 *Marginal compliance models*

We first modify 2 of the assumptions from Section 3.1 to take covariate information into account. We assume that randomization and monotonicity hold conditional on x, that is $Z \perp\!\!\!\perp \{Y_0, Y_1, A_0, A_1, X\}$ and $P(A_1 = 1|A_0 = 1, X = x) \geqslant P(A_1 = 1|A_0 = 0, X = x)$.

Let $X_z \subseteq X$ denote covariates that are predictive of compliance in arm $Z = z$ (i.e. $[A_z|X] = [A_z|X_z]$), $z = 0, 1$. We assume $\psi_z(X_z) = P(A_z = 1|X_z) = m_z(X_z; \lambda_z)$, where $m_z(X_z, \lambda_z)$ is a user-specified function indexed by a finite-dimensional parameter λ_z, $z = 0, 1$. For example, one could specify logistic regression models $\psi_z(X_z) = \text{logit}^{-1}(X_z^T \lambda_z)$, for $z = 0, 1$. Recall that $A = ZA_1 + (1 - Z)A_0$ is the observed compliance variable. By the randomization assumption, $[A_0|X] = [A|X, Z = 0]$ and $[A_1|X] = [A|X, Z = 1]$. Therefore, the λ parameters, and hence the marginal probabilities $\psi_0(X_0)$ and $\psi_1(X_1)$, are identifiable from the observed data. The usual diagnostic methods can be used to check the adequacy of the $m_z(\cdot; \cdot)$.

4.2 *Association model*

To complete the specification of the joint distribution $[A_0, A_1|X]$, we link the 2 marginal distributions together, using a model that satisfies the following: (1) any value of ϕ in the specified range should yield a model for $[A_0, A_1|X]$ that is fully compatible with the assumed marginal distributions $\psi_0(X_0)$ and $\psi_1(X_1)$ and (2) any value of ϕ in the specified range should lead to valid conditional probabilities (e.g. $P(A_1|A_0, x)$ must be between 0 and 1).

Given Assumptions 1–5 and marginal models $\psi_z(X_z)$, the conditional distribution $P(A_1 = 1|A_0 = 1, X = x)$ is bounded below by $\psi_1(x)$ and above by $U(x) = \min\{1, \frac{\psi_1(x)}{\psi_0(x)}\}$. We propose the following association model to complete the specification of $[A_0, A_1|X]$:

$$P(A_1 = 1|A_0 = 1, X = x) = \psi_1(x) + \phi\{U(x) - \psi_1(x)\}, \quad 0 \leqslant \phi \leqslant 1. \tag{4.1}$$

Interpretation of ϕ. If $\phi = 0$, then $P(A_1 = 1|A_0 = 1, X = x) = \psi_1(x)$, for all x (i.e. A_1 is independent of A_0 given X). If $\phi = 1$, then $P(A_1 = 1|A_0 = 1, x) = U(x)$, the largest probability compatible with the marginal distributions. For example, if more people with covariates x would comply with new treatment compared to standard treatment (i.e. $\psi_1(x) > \psi_0(x)$), then setting $\phi = 1$ implies that everyone with covariates x who would comply with the standard treatment would also comply with the new treatment. The marginal distributions $\psi_z(X_z)$ and conditional probability (4.1) imply a model for $P(A_1 = 1|A_0 = 0, X = x)$. Specifically, the probability $\psi_1(x) = P(A_1 = 1|X = x)$ can be written as

$$\begin{aligned}\psi_1(x) &= \sum_{a=0}^{1} P(A_1 = 1|A_0 = a, X = x)P(A_0 = a|X = x) \\ &= [\psi_1(x) + \phi\{U(x) - \psi_1(x)\}]\psi_0(x) + P(A_1 = 1|A_0 = 0, x)\{1 - \psi_0(x)\}.\end{aligned}$$

Solving for $P(A_1 = 1|A_0 = 0, X = x)$ gives

$$P(A_1 = 1|A_0 = 0, X = x) = \frac{\psi_1(x) - [\psi_1(x) + \phi\{U(x) - \psi_1(x)\}]\psi_0(x)}{1 - \psi_0(x)},$$

and therefore the association model generalizes to

$$P(A_1 = 1|A_0 = a, X = x) = a[\psi_1(x) + \phi\{U(x) - \psi_1(x)\}] + (1 - a)\frac{\psi_1(x) - [\psi_1(x) + \phi\{U(x) - \psi_1(x)\}]\psi_0(x)}{1 - \psi_0(x)}, \quad (4.2)$$

which shows the joint distribution $f(A_0, A_1|X) = f(A_0|X)f(A_1|A_0, X)$ is a function of the 2 marginal distributions $\psi_0(X)$ and $\psi_1(X)$ and the parameter ϕ.

5. INFERENCE METHODS

Inferences are based on an observed-data posterior distribution. The main strategy pursued here is to specify the full-data model and then impose necessary assumptions for identifying parameters of interest. Let $\beta = (\phi, \lambda_1, \lambda_0)$ and let θ denote the collection of parameters that characterize $[(Y_0, Y_1) \mid S]$. We denote the full-data joint distribution generically as $[Y_0, Y_1, A_0, A_1, Z, X]$ (equivalently, $[Y_0, Y_1, S, Z, X]$) and decompose it using the following factorization:

$$f(Y_0, Y_1, A_0, A_1, Z, X \mid \theta, \beta, \xi) = f(Y_0, Y_1 \mid S, Z, X, \theta) \times f(A_0, A_1 \mid Z, X, \beta) \times f(X, Z \mid \xi). \quad (5.1)$$

The variable S appears in the first factor because conditioning on (A_0, A_1) is equivalent to conditioning on S. The full-data likelihood contribution for a single individual is any function proportional in (θ, β, ξ) to the joint distribution (5.1), evaluated at $F_i = (Y_{0i}, Y_{1i}, A_{0i}, A_{1i}, Z_i, X_i)$, and is denoted generically by $L(\theta, \beta, \xi \mid F) = \prod_{i=1}^{n} L(\theta, \beta, \xi \mid F_i)$.

Because it is not possible to observe the full data, inference must be based on an observed-data posterior distribution. The likelihood part of the observed-data posterior is derived by integrating missing observations out of the joint distribution (5.1). Before proceeding, we make 2 assumptions about the full-data likelihood. First, we assume that the potential outcomes are jointly independent of Z and X within principal stratum so that $f(Y_0, Y_1 \mid S, Z, X, \theta) = f(Y_0, Y_1 \mid S, \theta)$. This implies that the causal effects within strata are independent of both Z and X. For discrete X, this assumption can be relaxed to the extent that this methodology can be applied separately at distinct levels of X. Second, by randomization, $f(A_0, A_1 \mid Z, X, \beta) = f(A_0, A_1 \mid X, \beta)$.

To parameterize the joint distribution of potential responses, observe that there are 4 possible realizations of (Y_0, Y_1) at each level of S. Therefore, $f_1(Y_0, Y_1 \mid S, \theta)$ can be parameterized in terms of the probabilities $\theta_{y_0 y_1}(s) = \text{pr}(Y_0 = y_0, Y_1 = y_1 \mid S = s)$, where $\sum_{y_0=0}^{1} \sum_{y_1=0}^{1} \theta_{y_0 y_1}(s) = 1$ for any s. The exclusion restriction implies that $\theta_{10}(0, 0) = \theta_{01}(0, 0) = 0$. Hence,

$$f(Y_0, Y_1 \mid S, \theta) = \{\theta_{00}(0, 0)^{(1-Y_1)(1-Y_0)}\theta_{11}(0, 0)^{Y_1 Y_0}\}^{I\{S=(0,0)\}} \times \prod_{s \in \{(0,1),(1,0),(1,1)\}} \{\theta_{00}(s)^{(1-Y_1)(1-Y_0)}\theta_{01}(s)^{(1-Y_0)Y_1}\theta_{10}(s)^{Y_0(1-Y_1)}\theta_{11}(s)^{Y_0 Y_1}\}^{I(S=s)}.$$

The model for potential compliance variables is specified using the factorization $f(A_0, A_1 \mid X) = f(A_0 \mid X)f(A_1 \mid A_0, X)$, where the first factor is the Bernoulli mass function $f(A_0 \mid X) = \psi_0(X)^{A_0}\{1 - \psi_0(X)\}^{1-A_0}$ and the second factor is Bernoulli with probability given by (4.2).

To complete the model specification, we assume that the parameters (θ, β) are *a priori* jointly independent of ξ; that is $p(\theta, \beta, \xi) = p(\theta, \beta)p(\xi)$ which obviates the need to specify $f(x, z|\xi)$. The

observed-data likelihood is obtained by integrating over the sample space of missing potential outcomes and can be written as

$$L(\pi, \beta \mid Y, A, Z, X) = \sum_{s \in \{(0,0),(0,1),(1,0),(1,1)\}} (\pi_Z^{(s)})^Y \{1 - \pi_Z^{(s)}\}^{1-Y} P(S = s \mid X, \beta) G(s, A, Z), \quad (5.2)$$

where

$$G(s, A, Z) = I\{s = (0, 0)\}(1 - A) + I\{s = (1, 0)\}\{A(1 - Z) + (1 - A)Z\}$$
$$+ I\{s = (0, 1)\}\{AZ + (1 - A)(1 - Z)\} + I\{s = (1, 1)\}AZ.$$

See Section B of the supplementary material available at *Biostatistics* online for detailed calculation. Specifics about priors are given in Section 6.

6. Analysis of CTQ data

6.1 *Data*

The CTQ study was described in Section 2. The outcome was defined as continuous abstinence during the course of the trial, a common end point in smoking cessation trials. We defined someone as compliant with their assigned intervention during the 4 weeks prior to the quit date if they attended at least 2 of the 3 required sessions every week and 3 sessions at least twice. Therefore, individuals with $S = (1, 1)$ could be thought of as those who would be highly compliant with either intervention during the pre-quit week phase of the trial.

6.2 *Model specification*

Compliance model. We considered numerous predictors of compliance at the model selection stage, including general characteristics of the subject (marital status, employment status, race/ethnicity, age, income), measures of nicotine dependence (Fagerstrom scale of nicotine dependence (Heatherton *and others*, 1991), number of cigarettes smoked per day, age started smoking), previous quit attempts (number of quit attempts, ever quit for 24 hours), and body weight–related variables (BMI, percent body fat, history of dieting). We selected a compliance model for each arm based on parsimony, predictive accuracy, and model fit. We decided on the following model for $\psi_0(x)$:

$$\text{logit}^{-1}(\lambda_{00} + \lambda_{01}\text{age} + \lambda_{02}\text{employ} + \lambda_{03}\text{married} + \lambda_{05}\text{employ} \times \text{married} + \lambda_{06}\text{ed12} + \lambda_{06}\text{ed15}),$$

where age is years of age of the subject at baseline, employ is an indicator variable for employment at baseline and married is an indicator variable that the subject is married, and ed12 and ed15 are indicators that the subject has at least 12 and 15 years of education, respectively. This model fitted the data well (Hosmer–Lemeshow goodness-of-fit statistic = 5.57 on 8 degrees of freedom; $p = 0.69$). For the exercise condition, we chose the model

$$\psi_1(x) = \text{logit}^{-1}(\lambda_{10} + \lambda_{11}\text{ed12} + \lambda_{12}\text{ed15}).$$

Covariates other than education level did not seemed to have an effect on the exercise compliance probability.

Gibbs sampler. We specified uniform (0, 1) priors for the $\pi_z^{(s)}$ parameters, $z = 0, 1$, $s = (0, 0), \ldots, (1, 1)$. Proper but relatively flat normal priors were assumed for the λ's with mean 0 and variance 1000. At each fixed value of ϕ, we generated 60 000 draws of the parameters using WinBUGS. The first 10 000 draws were discarded. Convergence appeared to be reached by about the 1000th draw based on trace plots and the Gelman–Rubin statistic.

6.3 *Results*

The value of ϕ affects the estimated number of subjects in each stratum (i.e. the size of each compliance-type subpopulation). In Figure 1, we display estimates of the proportion of subjects in each principal stratum at various values of ϕ. These estimates were calculated by generating a value of S from the posterior distribution for each subject at each Gibbs draw, calculating the proportion for each value of S, and then averaging those over all 50 000 draws. When ϕ is assumed to be near 0, it follows that there are a nontrivial number of subjects who would comply in one arm but not in the other (strata (1, 0) and (0, 1)). When ϕ is set near 1, nearly everyone who would comply in one arm would also comply in the other. Therefore, the estimated proportion of subjects in strata (0, 0) and (1, 1) increased as ϕ increased.

In Table 1, we present the 2.5, 50, and 97.5 percentiles of the posterior distributions of λ_0 and λ_1, under the assumption of conditional independence ($\phi = 0$); the results were similar at other values of ϕ. For

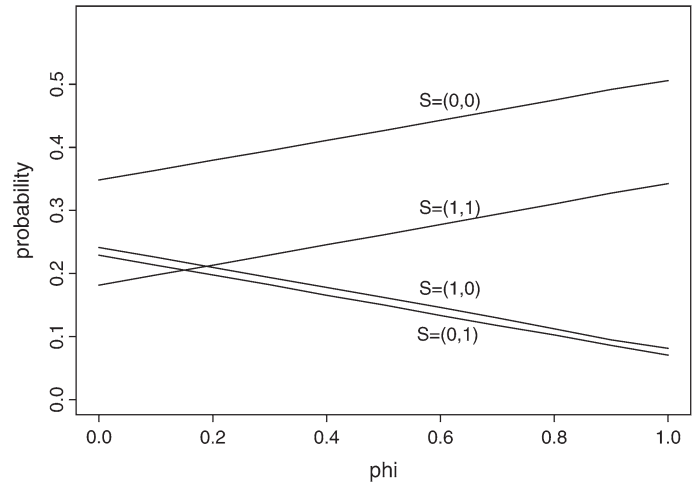

Fig. 1. Estimated proportion of population in each principal stratum at each value of ϕ.

Table 1. *Estimates of the* 2.5, 50, *and* 97.5 *percentiles of the posterior distribution of the compliance model parameters when* $\phi = 0$

Parameter	2.5	50	97.5
Contact compliance model			
Intercept	−4.36	−2.33	−0.50
Age	0.02	0.06	0.10
Employed	−1.15	0.01	1.21
Married	−0.02	1.31	2.73
Married × employed	−3.15	−1.56	−0.02
Education $\leqslant$ 12 years	−0.67	0.19	1.04
Education $\leqslant$ 15 years	−1.74	−0.81	0.07
Exercise compliance model			
Intercept	−0.21	−0.49	1.21
Education $\leqslant$ 12 years	−1.22	−0.41	0.39
Education $\leqslant$ 15 years	−1.82	−0.89	0.01

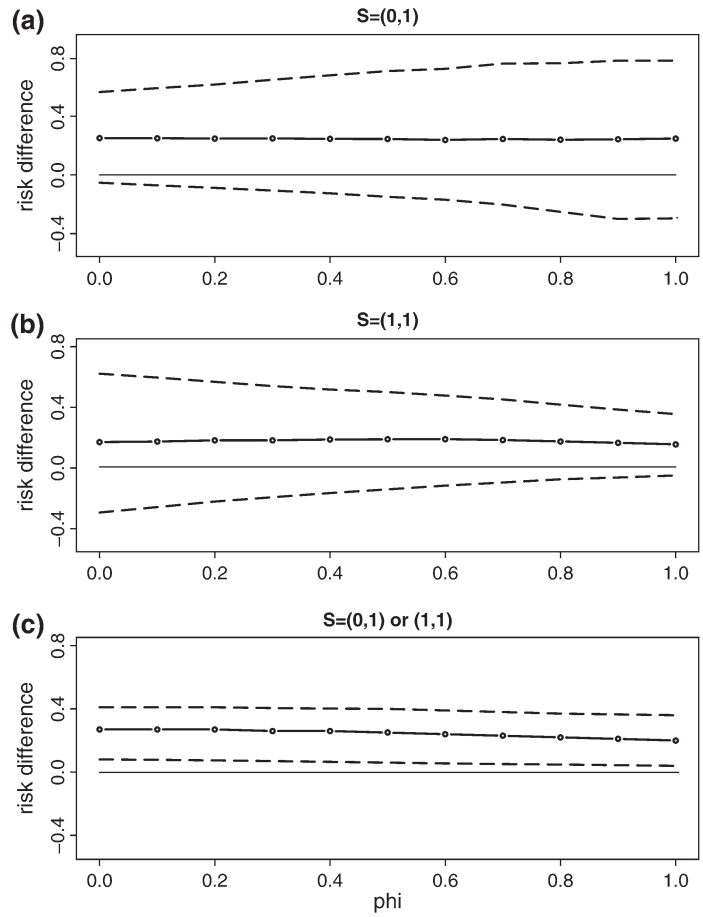

Fig. 2. Plot of 2.5, 50, and 97.5 percentiles of posterior distribution of the risk difference for stratum (0, 1), $\pi_1^{(0,1)} - \pi_0^{(0,1)}$, in (a); for stratum (1, 1), $\pi_1^{(1,1)} - \pi_0^{(1,1)}$, in (b); and the combined stratum (0, 1) and (1, 1), $\pi_1^{(0,1)\cup(1,1)} - \pi_0^{(0,1)\cup(1,1)}$, in (c), at different values of ϕ.

the control arm compliance model, older subjects and those with more education appeared more likely to comply. In addition, subjects who were married and unemployed were more likely to comply than subjects who were married and employed or who were not married. Based on results from the exercise compliance model, subjects with 16 or more years of education appeared more likely to comply than subjects who had fewer years of education.

Summary statistics relating to the posterior distribution of $\pi_z^{(s)}$ are provided in Section D of the supplementary material available at *Biostatistics* online.

Causal effect of exercise among compliers. Primary interest is in $\pi_1^{(1,1)} - \pi_0^{(1,1)}$, the difference in cessation rates between exercise and contact groups for the subgroup that would comply with either intervention. The posterior median, along with 95% credible interval for this causal risk difference, is displayed in Figure 2(b) at various values of ϕ. For all values of ϕ, the posterior median is positive, indicating that among highly compliant subjects, the cessation probability is higher when randomized to the exercise condition. For values of ϕ near 0, the 95% interval is very wide, reflecting lack of information about the intervention effect. A value of ϕ between 0.5 and 1 might be more plausible, however, as the

2 interventions are similar in terms of what is expected from participants (both require attending a class 3 days per week). For large values of ϕ, the interval becomes much more narrow because class membership in group 3 increases (see Figure 1). The interval overlaps the null value of 0, even when $\phi = 1$, but most of the posterior mass is greater than 0. The posterior median is 0.16 with 95% interval $(-0.04, 0.36)$ at $\phi = 1$.

ITT effect for entire population. The traditional ITT analysis can be thought of as the causal effect of Z on the entire population. The sensitivity parameter ϕ has no impact here because varying ϕ does not shift the population in $S \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. The posterior median and 95% credible interval for the $E(Y_1 - Y_0)$ were 0.085 and (0.002, 0.170), suggesting a benefit of being randomized to the exercise group instead of the wellness group.

Causal contrasts for other subpopulations. The model can be used to infer other contrasts of potential interest, such as the effect of exercise among subjects who would comply with exercise but not with wellness. Figure 2(a) displays posterior quantiles of $\pi_1^{(0,1)} - \pi_0^{(0,1)}$ as a function of ϕ. As shown in Figure 1, when ϕ was near 0, group membership in $S = (0, 1)$ was at its largest. Therefore, as ϕ increased, the credible intervals for the causal effect in arm $S = (0, 1)$ became wider. The majority of the posterior draws of causal risk difference were greater than 0, suggesting a benefit of exercise for this subpopulation. However, the 95% credible intervals always included 0, so we cannot rule out lack of effect.

Investigators might also be interested in the subpopulation that would comply with exercise if offered, regardless of whether or not they would comply with the contact intervention (i.e. $S = (0, 1)$ or $(1, 1)$). In Figure 2(c), we display the posterior median and 95% credible interval for the ITT effect in the $S = (0, 1)$ or $(1, 1)$ subpopulation at various values of ϕ. The interval did not overlap with 0 at any value of ϕ, indicating a benefit of exercise for this subpopulation. The interval is much less sensitive to values of ϕ than with the population of compliers $S = (1, 1)$ or exercise-only compliers $S = (0, 1)$. With $S = (1, 1)$, the width of the 95% interval for $\pi_1^{(1,1)} - \pi_0^{(1,1)}$ changed dramatically as a function of ϕ because the estimated proportion with $S = (1, 1)$ increases sharply with ϕ (Figure 1); however, the proportion of subjects in $S = (0, 1)$ or $(1, 1)$ is nearly constant as a function of ϕ. As ϕ increases, there is a shift from $S = (0, 1)$ to $S = (1, 1)$, but the combined population remained about the same. Because $S = (0, 1)$ or $(1, 1)$ includes everyone who would comply with exercise, information about that class comes from the marginal distribution of A_1.

Characteristics of subjects in each stratum. The model proposed here can also be used to characterize the covariate distribution of individuals in each principal stratum; i.e. $[X \mid S]$, as opposed to $[S \mid X]$, which is summarized in Table 1. This can be useful for characterizing the subpopulation of compliers in terms of their covariate profile. Detailed results are given in Section D of the supplementary material available at *Biostatistics* online.

The impact of covariates. We explored the impact of modeling compliance as a function of covariates by fitting a model with no covariates in the compliance models. As described previously, without covariates the strata probabilities and causal effects are not point identified in this setting (2 active treatments). In the absence of priors, however, these parameters can be bounded. Figure 3 includes plots of the estimated posterior densities of $\pi_0^{(1,1)}$, $\pi_1^{(1,1)}$, and $\pi_1^{(1,1)} - \pi_0^{(1,1)}$ in the models with and without covariates, for ϕ equal to 0 and 1. When $\phi = 0$, the posterior distributions from the model without covariates had flat sections at the maximum height of the posterior density, reflecting the lack of information needed to identify a maximum point. Instead, there were a range of values that could be viewed as "most likely." The plots from the model with covariates showed distributions that were more narrow and had clear maxima. For ϕ equal to 1, as expected, the impact of covariates was less pronounced because a subject who complied with one intervention would likely have complied with the other. Observed compliance is essentially all that is needed to identify the stratum in that case. But because ϕ is unknown, Figure 3 demonstrates how conditioning on covariates in the compliance models can reduce the range of likely values of the parameters of interest over the range of plausible values of ϕ.

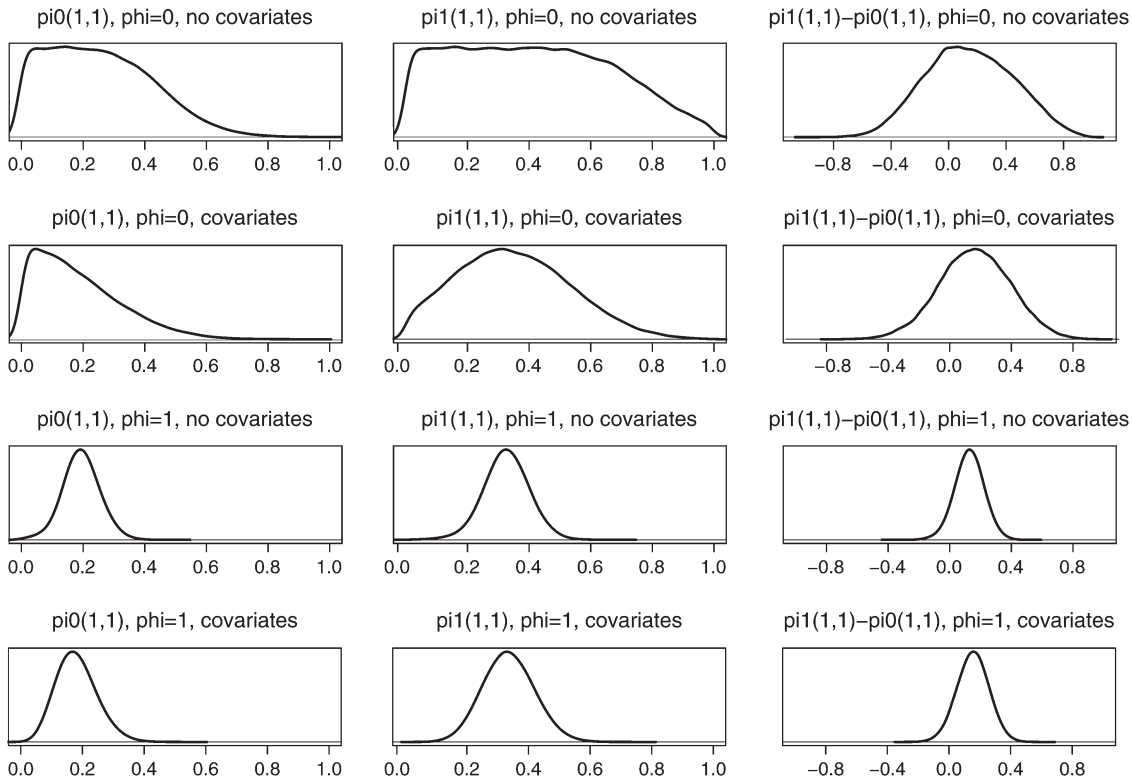

Fig. 3. Plot of posterior distribution of $\pi_0^{(1,1)}$, $\pi_1^{(1,1)}$, and $\pi_1^{(1,1)} - \pi_0^{(1,1)}$ for the models with and without covariates and at $\phi = 0$ and $\phi = 1$.

Sensitivity to Assumption 5. While we believe Assumption 5 is plausible, we also investigate the sensitivity of inference to that assumption. Without Assumption 5, the lower bound of the association model $P(A_1 = 1|A_0 = a, X = x)$ is

$$P(A_1 = 1|A_0 = a, X = x) = aL(x) + (1 - a)\frac{\psi_1(x) - L(x)\psi_0(x)}{1 - \psi_0(x)},$$

where $L(x) = \max[0, \{\psi_1(x) + \psi_0(x) - 1\}/\psi_0(x)]$. We fitted the full model assuming this is the true association model. The posterior median and 95% credible interval for $\pi_1^{(1,1)} - \pi_0^{(1,1)}$ were 0.10 and $(-0.55, 0.69)$, substantially wider compared to using Assumption 5 with $\phi = 0$. The posterior mean for the proportion of subjects in stratum $S = (1, 1)$ was 0.05. Because the compliance rates are similar in the 2 arms, if we do not make Assumption 5, then we allow for the possibility that nearly everyone who would comply with one treatment would not comply with the other. In that scenario, strata $S = (1, 0)$ and $S = (0, 1)$ are the largest. This demonstrates both the benefit and the risk of Assumption 5: if correct, it can substantially shrink the bounds on the causal effects, if not it can produce misleading results.

7. Discussion

Experimental studies with 2 active treatments are fairly common, especially in behavioral intervention studies. When there is noncompliance in each arm, drawing causal comparisons between the 2

interventions is challenging because the population of subjects who would comply with one intervention might differ from the population that would comply with the other. We have proposed a model that uses compliance-predictive covariates to identify principal effects (causal contrasts in subpopulations defined by compliance behavior) up to a sensitivity parameter. Identification of the principal strata probabilities relies on the specification of 2 marginal compliance models, conditional on covariates, and an association model that ensures that the joint probability remains within the bounds of the parameter space. At each value of the association parameter, we are able to make inference about the principal effects. In the smoking cessation analysis, the evidence suggested that the exercise regimen would yield better results than the wellness therapy, among subjects who would comply with either intervention, although the results were not conclusive.

Our findings suggest that researchers who are planning trials with 2 active treatments would benefit from collecting compliance information from both arms, as well as baseline covariate information that they expect will be predictive of compliance. By collecting this additional information, the methods described in this paper can be used to identify causal effects among subpopulations defined by their compliance pattern. This should provide investigators with a far greater understanding of the effects of the 2 treatments, beyond what could be learned from the usual ITT approach and beyond what can be learned without using covariates.

The proposed methodology has other applications. For example, if A is a binary mediating variable, our approach could be used to estimate direct effects of the intervention on the outcome among subpopulations for whom the intervention has no effect on the mediator (i.e. the population $A_0 = A_1$; Mealli and Rubin, 2003).

As pointed out by a reviewer, the use of covariates might cause concerns similar to those in case–control studies (e.g. concerns about overmatching). Our view is that if covariates that are predictive of compliance are available, including them in the model accomplishes 2 things. First, it can reduce the width of the credible intervals. Second, it can help to answer an important question that behavioral researchers want more information about, who are the compliers. With that said, there is some risk involved with introducing covariates into the model (covariate selection, model specification). However, the fit of the marginal part of the compliance models can be checked using standard regression diagnostic methods.

A potential limitation with our analysis is that the SUTVA assumption might not hold. The intervention was typically held in small groups, and it is possible that group dynamic could affect compliance and/or cessation status. Based on our conversations with CTQ investigators, we believe that substantial interference between subjects was unlikely. However, we cannot rule it out.

ACKNOWLEDGMENTS

We thank the editor and associate editor for their insightful comments and suggestions. *Conflict of Interest:* None declared.

FUNDING

The National Institutes of Health (R01-HL-79457, R29CA59660, K07CA01757).

REFERENCES

ANGRIST, J., IMBENS, G. AND RUBIN, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

BALKE, A. AND PEARL, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* **92**, 1171–1176.

CHENG, J. AND SMALL, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society, Series B* **68**, 815–836.

DOMINICI, F., ZEGER, S. L., PARMIGIANI, G., KATZ, J. AND CHRISTIAN, P. (2006). Estimating percentile-specific effects in counterfactual models: a case study of micronutrient supplementation, birth weight, and infant mortality. *Journal of the Royal Statistical Society, Series C* **55**, 1–20.

ESCOBEDO, L. G. AND PEDDICORD, J. P. (1996). Smoking prevalence in US birth cohorts: the influence of gender and education. *American Journal of Public Health* **86**, 231–236.

FRANGAKIS, C. AND RUBIN, D. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

GOETGHEBEUR, E. AND LAPP, K. (1997). The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Applied Statistics* **46**, 351–364.

HEAGERTY, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.

HEATHERTON, R. F., KOZLOWSKI, L. T., FRECKER, R. C. AND FAGERSTROM, K. O. (1991). The Fagerstrom test for nicotine dependence: a revision of the Fagerstrom tolerance questionnaire. *British Journal of Addiction* **86**, 1119–1127.

HIRANO, K., IMBENS, G., RUBIN, D. AND ZHOU, X. (2000). Assessing the effect of an influenza vaccine in an encouragement design with covariates. *Biostatistics* **1**, 69–88.

JOFFE, M. (2001). Using information on realized effects to determine prospective causal effects. *Journal of the Royal Statistical Society, Series B* **63**, 759–774.

MANSKI, C. (1990). Non-parametric bounds on treatment effects. *American Economic Review, Papers and Proceedings* **80**, 319–323.

MARCUS, B. H., ALBRECHT, A. E., KING, T. K., PARISI, A. F., PINTO, B. M., ROBERTS, M., NIAURA, R. AND ABRAMS, D. B. (1999). The efficacy of exercise as an aid for smoking cessation in women. *Archives of Internal Medicine* **159**, 1229–1234.

MARCUS, B. H., LEWIS, B., HOGAN, J. W., KING, T. K., ALBRECHT, A. E., BOCK, B. AND PARISI, A. F. (2005). The efficacy of moderate-intensity exercise as an aid for smoking cessation in women: a randomized controlled trial. *Nicotine and Tobacco Research* **7**, 871–880.

MEALLI, F. AND RUBIN, D. B. (2003). Commentary: Assumptions allowing the estimation of direct causal effects. *Journal of Econometrics* **112**, 79–87.

ROBINS, J. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest, L., Freedoman, H. and Bailey, A. (editors), *Health Services Research Methodology: A Focus on AIDS*. Washington, DC: U.S. Public Health Service, National Center for Health Services Research, pp. 113–159.

ROBINS, J. M. (1994). Correcting for non-compliance in randomised trials using structural nested mean models. *Communications in Statistics* **23**, 2379–2412.

ROBINS, J. AND ROTZNITZKY, A. (2004). Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* **91**, 763–783.

[*Received November 2, 2006; revised June 20, 2007; accepted for publication July 10, 2007*]