*Collection of Biostatistics Research Archive*
COBRA Preprint Series

*Year* 2006 *Paper* 6

# Causal Comparisons in Randomized Trials of Two Active Treatments: The Effect of Supervised Exercise to Promote Smoking Cessation

Jason Roy* Joseph W. Hogan†

*University of Rochester, jaroy@geisinger.edu

†Brown University, jwh@brown.edu

# Causal Comparisons in Randomized Trials of Two Active Treatments: The Effect of Supervised Exercise to Promote Smoking Cessation

Jason Roy and Joseph W. Hogan

**Abstract**

In behavioral medicine trials, such as smoking cessation trials, two or more active treatments are often compared. Noncompliance by some subjects with their assigned treatment poses a challenge to the data analyst. Causal parameters of interest might include those defined by subpopulations based on their potential compliance status under each assignment, using the principal stratification framework (e.g., causal effect of new therapy compared to standard therapy among subjects that would comply with either intervention). Even if subjects in one arm do not have access to the other treatment(s), the causal effect of each treatment typically can only be identified from the outcome, randomization and compliance data within certain bounds. We propose to use additional information – compliance-predictive covariates – to help identify the causal effects. Our approach is to specify marginal compliance models conditional on covariates within each arm of the study. Parameters from these models can be identified from the data. We then link the two compliance models through an association model that depends on a parameter that is not identifiable, but has a meaningful interpretation; this parameter forms the basis for a sensitivity analysis. We demonstrate the benefit of utilizing covariate information in both a simulation study and in an analysis of data from a smoking cessation trial.

# Causal comparisons in randomized trials of two active treatments: The effect of supervised exercise to promote smoking cessation

**Summary.** In behavioral medicine trials, such as smoking cessation trials, two or more active treatments are often compared. Noncompliance by some subjects with their assigned treatment poses a challenge to the data analyst. Causal parameters of interest might include those defined by subpopulations based on their potential compliance status under each assignment, using the principal stratification framework (e.g., causal effect of new therapy compared to standard therapy among subjects that would comply with either intervention). Even if subjects in one arm do not have access to the other treatment(s), the causal effect of each treatment typically can only be identified from the outcome, randomization and compliance data within certain bounds. We propose to use additional information – compliance-predictive covariates – to help identify the causal effects. Our approach is to specify marginal compliance models conditional on covariates within each arm of the study. Parameters from these models can be identified from the data. We then link the two compliance models through an association model that depends on a parameter that is not identifiable, but has a meaningful interpretation; this parameter forms the basis for a sensitivity analysis. We demonstrate the benefit of utilizing covariate information in both a simulation study and in an analysis of data from a smoking cessation trial.

Key Words: Bounds; Causal effect; Latent class model; Noncompliance; Principal stratification; Potential outcomes; Randomized trial.

# 1 Introduction

In many clinical trials, the objective is to compare a new treatment to placebo or no treatment. In trials of new behavioral interventions, however, the experimental design frequently calls for comparing a new or improved intervention with the current standard of care – an 'active' treatment – rather than a placebo. For example, in trials of interventions designed to promote behaviors such as smoking cessation, weight loss, or improved adherence to medication, new interventions typically are compared to a standard of care such as cognitive behavioral therapy.

In a trial of active treatment versus placebo, interest typically focuses on the effect of the treatment relative to no treatment. Causal comparisons might be defined in terms of outcome under full compliance with treatment versus no compliance with treatment. For trials with two active treatments, however, it is possible to contemplate more than one treatment effect, such as new treatment versus standard of care, new treatment versus no treatment, or new treatment versus not having new treatment (in this last case, 'not having new treatment' may be defined as any treatment not involving the new one).

A key complication in drawing inference about causal effects in any trial is that compliance is rarely perfect. In trials of an active treatment versus placebo, it is possible to recover causal effects under some reasonably mild assumptions. For all-or-none compliance situations (i.e., compliance is a binary variable), the method of instrumental variables can be particularly useful (Angrist, Imbens and Rubin, 1996). Causal effects among compliers (subjects that would take treatment if offered) are identifiable if there are no defiers (subject who would take the active treatment in the control arm but not in the treatment arm). This assumption is particularly reasonable in placebo controlled trials, or 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 Rotnitzky 2004). Structural mean models can also be used when compliance is continuous, and if there are interactions between the causal effect and 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; Balke and Pearl 1997; Joffe 2001).

In other settings, particularly when a standard of care exists and must be included for comparison, the trial will compare two or more active treatments. For example, in trials of new antiviral therapy regimens for those with HIV infection, comparison with a placebo is typically not ethically viable. Trials of behavioral interventions, where new interventions are usually tested relative to a standard or existing behavioral or pharmacologic therapy, also are representative. A specific example is the Commit to Quit trials (Marcus et al. 1999), comprising two longitudinal follow up studies of supervised exercise to promote smoking cessation. Each study had two treatment arms. All participants received cognitive-behavioral smoking cessation therapy (CBT). For those in the intervention arm, CBT was augmented by an individualized, supervised exercise program. In order to equalize increased contact hours between the two 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' and standard therapy augmented by an exercise regimen. Another particularly useful feature of behavioral intervention trials for drawing causal inferences is that compliance is fully observed when the intervention comes in the form of group therapy or other supervised activity.

A useful framework for drawing causal inference about the effect of a post-randomization outcome such as compliance with treatment is principal stratification (Frangakis and Rubin 2002). Causal effects are characterized within subpopulations, or principal strata, defined by potential values of a post-treatment stratification variable. When compliance is the post-treatment variable, principal stratification can be used to identify causal contrasts associated with receiving treatments under study. For example, in a trial where individuals are randomized to one of two active treatments, say $A$ and $B$, and where compliance is a binary post-treatment variable (1 = yes, 0 = no), there are four principal strata defined by the pairs of potential values of compliance, those who

would: (i) comply with both $A$ and $B$, (ii) comply with $A$ but not $B$, (iii) comply with $B$ but not $A$, and (iv) comply with neither.

The causal effect of *being randomized* to one treatment arm versus another is the intention to treat (ITT) effect. It is estimable for a population and, by virtue of randomization, the estimate captures a causal effect (here, the causal effect of being randomized to $A$ versus $B$, averaged over post-stratification variables). Causal effects of *receiving* treatment can be defined as (or in terms of) ITT contrasts within each principal stratum. For example, the causal effect of receiving $A$ versus $B$ is the effect of being randomized to $A$ versus $B$ *among those who would comply with either*; i.e. among those in stratum (i).

The ITT effect within other strata are still causal effects, but may not be of direct interest. For example, the ITT effect among those in stratum (ii) — comprising subjects who would comply with treatment $A$ but not treatment $B$ — gives the causal effect of receiving treatment A relative to receiving neither $A$ nor $B$.

Although the principal stratification approach is useful for defining the causal parameters of interest, there is still an identifiability problem. In the absence of covariates, one can derive bounds on causal contrasts associated with receiving one treatment relative to another. Cheng and Small (2006) derived bounds on the distribution of principal strata and causal effects for trials with two active treatments and non-compliance; how tight the bounds are depends on the assumptions one is willing to make.

The goal of our approach is to identify causal parameters of interest by utilizing information from baseline covariates that are predictive of compliance. Each individual has two potential values of the post-randomization variable, say $A_0$ and $A_1$, where $A_z$ is the indicator that the subject would comply with treatment $z$ for $z = 0, 1$; only one is observed. Our approach relies on a model for the joint distribution of $(A_0, A_1)$ as a function of baseline covariates $X$. The joint distribution is written in terms of the treatment-arm-specific means, together with an association model; the association is parameterized so as not to affect the marginal distributions and to ensure that the joint probabilities

stay within the appropriate bounds. Using the model for $[A_0, A_1 \mid X]$, and assuming it is correctly specified, we are able to identify probability of membership in each principal stratum up to a sensitivity parameter that characterizes within-subject association between compliance with each of the active treatments; e.g. conditional on baseline covariates, the probability that an individual would comply with treatment $A$, given compliance status for treatment $B$. We can then identify the causal parameters at each fixed value of the sensitivity parameter.

In Section 2 we describe the smoking cessation trial. We introduce the notation, assumptions and bounds on the causal effects in Section 3. In Section 4 we present the compliance models. The likelihood, prior and posterior distributions are described in Section 5. Section 6 describes a brief simulation study. Data from the CTQ smoking cessation trial are analyzed in Section 7. Finally, there are concluding remarks in Section 8.

## 2 Data from 'Commit to Quit' trial

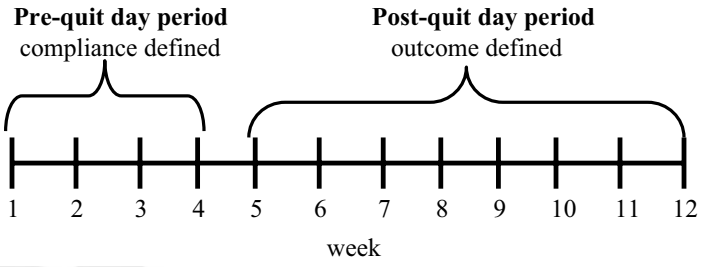
The Commit to Quit study (Marcus et al., 1999) is a randomized controlled trial designed to assess the efficacy of supervised vigorous exercise as an adjuvant to cognitive behavioral therapy to promote smoking cessation among women. A primary motivation behind studying women exclusively is that in the mid-1990's, smoking prevalence rates among women were declining at a slower rate than among men (Escobedo and Peddicord 1996). The study enrolled and randomized 281 women to receive CBT plus vigorous exercise (the new treatment) or 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 two 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 three times weekly, and each session was supervised by an exercise specialist. Women in the control arm participated in a program of supervised lectures, films and discussions three 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.

Smoking status was evaluated weekly over the course of 12 weeks. The target quit date was week 5 following randomization. Weekly quit status was assessed by self report and verified by carbon monoxide (<8ppm) and saliva cotinine (<10 ng/mL). To be considered abstinent, an individual needed to submit to testing and meet both the carbon monoxide and saliva cotinine criteria. The primary outcome of the study, and the one we consider here, was continuous abstinence during the 8 weeks after the quit date. Consistent with the criteria used in Marcus et al. (1999), an individual who was not present for scheduled testing for one or more occasions during the study could not be counted as continuously abstinent.

Subjects were expected to attend 3 wellness or exercise classes per week. We defined compliance based on the number of sessions actually attended during the first 4 weeks of the trial (i.e., during the weeks prior to the quit date). Specifically, we defined a subject as compliant with their assigned treatment if they attended two-thirds of the classes every week and attended all 3 classes at least 2 of the weeks, during weeks 1 to 4. For clarity, this is depicted in Figure 1. Compliance status was observed for all subjects (no missing data).

Figure 1: Commit to Quit trial




A total of 147 and 134 subjects were randomized to the wellness and exercise arms, respectively. Compliance was similar in the wellness and exercise arms (41% and 43%, respectively). A total of 10.8% of subjects in the wellness arm were continually abstinent during the 8 weeks after the quit date, compared to 19.4% in the exercise arm.

# 3 Principal stratification with two active treatments

## 3.1 Notation and assumptions

We consider experimental trials with two arms – each with an active treatment. Our approach could easily be extended to three arm trials with two active treatments (where one arm is a control group), such as considered by Cheng and Small (2006). Let $Z_i \in \{0, 1\}$ denote randomization indicator for subject $i$ ($i = 1, \cdots, n$), where $Z_i = 1$ indicates randomization to the new treatment (e.g., supervised exercise plus CBT), and $Z_i = 0$ indicates randomization to standard therapy (e.g., wellness sessions plus CBT). For each individual, let $A_z \in \{0, 1\}$ denote compliance with *assigned* treatment under treatment assignment $z$. Therefore, each person has two potential treatments $A_0$ and $A_1$ that characterize their compliance behavior under either treatment assignment; however only $A = ZA_1 + (1 - Z)A_0$ is observed. Similarly, each subject also has two potential outcomes, $Y_0$ and $Y_1$, but only one is observed; we denote this as $Y = ZY_1 + (1 - Z)Y_0$. We next describe several key assumptions that we will rely on in the development of analytic methods.

*Assumption 1. Stable Unit Treatment Value Assumption (SUTVA) (Angrist, Imbens and Rubin 1996).* The potential compliance values of a subject do not depend on the treatment assignment of other subjects. In addition, the potential outcomes for a given subject do not depend on the assigned treatments or compliance status of other subjects.

*Assumption 2. Randomization.* We assume treatment is randomly assigned, which implies $Z \perp\!\!\!\perp \{Y_0, Y_1, A_0, A_1\}$.

*Assumption 3. Exclusion restriction.* Treatment assignment $Z$ only affects the outcome through its effect on treatment received.

Assumptions 1-3 imply that $Z$ is an instrumental variable.

*Assumption 4. Treatment access restriction.* Subjects in group $Z = z$ do not have access to the active treatment assigned in arm $Z = 1 - z$, for $z = 0, 1$. In the CTQ study, this assumption holds because subjects in the wellness arm were not allowed to attend the exercise classes, while subjects in the exercise arm were not allowed to attend the wellness classes.

*Assumption 5. Monotonicity.* Compared with the general population, the subpopulation who would comply with active control ($Z = 0$) are at least as likely to comply with the new intervention ($Z = 1$); i.e., $P(A_1 = 1|A_0 = 1) \geq P(A_1 = 1)$.

### 3.2 Defining causal effects

We assume each individual belongs to one of four basic principal strata (Frangakis and Rubin 2002) defined by unique combinations of $(A_0, A_1)$. In our case, these are $\{(0,0),(0,1),(1,0),(1,1)\}$; the random variable $S \in \{0,1,2,3\}$ denotes stratum membership. Hence the subpopulation with $S = 3$ characterizes those individuals who would comply with whichever treatment was offered; those with $S = 2$ would comply with the new treatment, but not the standard treatment; those with $S = 1$ would comply with standard treatment, but not the new treatment; and finally those with $S = 0$ would not comply with either intervention.

We are now ready to define causal contrasts of interest. In the most general terms, we are seeking the joint distribution $[(Y_0, Y_1) \mid S = s]$, which characterizes the causal effect of randomization to treatment within the subpopulation comprising principal stratum $s$. For example, the distribution $[(Y_0, Y_1) \mid S = 3]$ captures causal effect new therapy relative to standard therapy, among the subpopulation that would comply with either intervention.

Let $\pi_z(s) = P(Y_z = 1|S = s)$. One potential parameter of of interest is $\pi_1(3) - \pi_0(3)$, the causal risk difference among compliers. An odds ratio or risk ratio also could be used. In our analysis of the CTQ trial, we will use the risk difference $\pi_1(3) - \pi_0(3)$ as our primary target for inference. We also will consider risk differences derived from $[(Y_0, Y_1) \mid S \in \{2,3\}]$, which captures the effect of new treatment among subjects that would comply with new treatment; and from $[(Y_0, Y_1) \mid S = 2]$, which characterizes the effect of new therapy relative to no intervention among the subpopulation of subjects that would comply with the new therapy but not standard treatment.

### 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). However, one can derive bounds for the strata probabilities and the causal effects. First, consider bounds on the distribution of $S$. Using assumptions 1-4, Cheng and Small (2006) derived bounds for $P(S = s)$, $s = 0, 1, 2, 3$. They also derived bounds based on a fifth assumption, that $P(A_1 = 1|A_0 = 1) = 1$. That assumption is not realistic for the CTQ study, as it is not hard to imagine that there are subjects who would comply with exercise but not wellness, or vice versa. We instead make a weaker fifth assumption, that $P(A_1 = 1|A_0 = 1) \geq P(A_1 = 1)$. The bounds on the distribution of $S$ based on our assumptions 1-5 are as follows. Let $\psi_0 = P(A_0 = 1)$ and $\psi_1 = P(A_1 = 1)$. The bounds are then

$$\begin{aligned}
(1-\psi_1)(1-\psi_0) \leq \quad & P(S=0) \quad \leq \min(1-\psi_0, 1-\psi_1) \\
\max(0, \psi_0-\psi_1) \leq \quad & P(S=1) \quad \leq \psi_0(1-\psi_1) \\
\max(0, \psi_1-\psi_0) \leq \quad & P(S=2) \quad \leq \psi_1(1-\psi_0) \\
\psi_0\psi_1 \leq \quad & P(S=3) \quad \leq \min(\psi_0, \psi_1).
\end{aligned}$$

We next consider bounds on the causal effects. We focus here on the average causal effect of new treatment compared to standard treatment, among those subjects that would comply with either intervention, i.e., $\pi_1(3) - \pi_0(3)$. Let $P(S = 3) = p_3$ and $I = [\psi_0\psi_1, \min(\psi_0, \psi_1)]$, then, as derived in Cheng and Small (2006), the lower bound of $\pi_1(3) - \pi_0(3)$ is

$$\min_{p_3 \in I} \left\{ \max\left(0, 1 - \frac{1 - P(Y=1|Z=1, A=1)}{p_3/P(A=1|Z=1)}\right) - \min\left(1, \frac{P(Y=1|Z=0, A=1)}{p_3/P(A=1|Z=0)}\right) \right\}$$

and the upper bound is

$$\max_{p_3 \in I} \left\{ \min\left(1, \frac{P(Y=1|Z=1, A=1)}{p_3/P(A=1|Z=1)}\right) - \max\left(0, 1 - \frac{1 - P(Y=1|Z=0, A=1)}{p_3/P(A=1|Z=0)}\right) \right\}.$$

These bounds can be estimated, as $P(Y|Z, A)$ and $P(A|Z)$ involve only the observed data.

Our strategy, however, is to identify the causal effects by utilizing information from baseline covariates that are predictive of compliance. As a hypothetical example, suppose women with a low body mass index (BMI) were more likely to comply with the wellness intervention than were women with a high BMI (we could know that by using data from the $Z = 0$ arm). It is logical to believe that a woman with low BMI in arm $Z = 1$ who complied with exercise is more likely to be in stratum $S = 3$ than a woman who with high BMI who complied with exercise in arm $Z = 1$. That is the kind of information we plan to utilize to identify the causal effects. In the next section, we describe our strategy for modeling the joint distribution of $(A_0, A_1)$ conditional on baseline covariates.

# 4 Utilization of compliance-predictive covariates

Assume for each subject we have several baseline covariates $X$ that are associated with compliance in one or both arms. Our objective is to use the covariate information to model $[A_0, A_1|X]$; however, only one of $A_0$ and $A_1$ are observed for each subject. Therefore we are able (using randomization of $Z$) to identify the marginals $[A_0|X]$ and $[A_1|X]$, we cannot identify the conditional association between $A_0$ and $A_1$ given $X$. Our approach is to separately specify the marginal distributions $[A_0|X]$ and $[A_1|X]$, and to use a single nonidentifiable parameter $\phi$ to capture the association structure. This parameter will be used as part of a sensitivity analysis. This idea of specifying the two marginal distributions separately from the association model is related to work by Heagerty (2002), except here the association parameter is not identifiable.

## 4.1 Marginal compliance models

We first modify two of the assumptions from Section 3.1 to take covariate information into account. The randomization assumption now can be written as $Z \perp\!\!\!\perp \{Y_0, Y_1, A_0, A_1, X\}$. That is, we assume that the assignment mechanism is independent of potential outcomes, potential compliance status, and baseline covariates. We also assume that assumption 5 holds within levels of the covariates, i.e., $P(A_1 = 1|A_0 = 1, X = x) \geq P(A_1 = 1|X = x)$, for all $x$.

We next specify models for the marginal means of the potential compliance outcomes. Denote by $X_0 \subseteq X$ the covariates that are predictive of compliance in arm $Z = 0$ (i.e., $[A_0|X] = [A_0|X_0]$), and by $X_1 \subseteq X$ the covariates that are predictive of compliance in arm $Z = 1$. We would like to estimate these marginal distributions, but for identifiability, will typically need to define models for them. Specifically, we assume,

$$\psi_0(X_0) = P(A_0 = 1|X_0) = m_0(X_0; \lambda_0) \tag{1}$$

and

$$\psi_1(X_1) = P(A_1 = 1|X_1) = m_1(X_1; \lambda_1), \tag{2}$$

where $m_z(X_z, \lambda_z)$ are user-specified functions indexed by a finite-dimensional parameter $\lambda_z$, $z = 0, 1$. For example, one could specify logistic regression models $\psi_0(X_0) = \text{logit}^{-1}(X_0^T\lambda_0)$ and $\psi_1(X_1) = \text{logit}^{-1}(X_1^T\lambda_1)$. Recall that $A = ZA_1 + (1 - Z)A_0$ is the observed compliance variable. By Assumption 2 (randomization), $[A_0|X] = [A|X, Z = 0]$ and $[A_1|X] = [A|X, Z = 1]$. Therefore, the $\lambda$ 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 specification of the joint distribution $[A_0, A_1|X]$, we need to link the two marginal distributions together. Our goal is to specify a model for the association between $A_0$ and $A_1$ given $X$, that satisfies the following:

1. The association can be characterized by a single parameter $\phi$, which cannot generally be identified from the compliance data and covariates.

2. Any value of $\phi$ in the specified range should yield a model for $[A_0, A_1|X]$ that is fully compatible with the assumed marginal distributions (1) and (2).

3. Any value of $\phi$ in the specified range should lead to valid conditional probabilities (e.g., $P(A_1|A_0, x)$ must be between 0 and 1).

Our association model must account for the fact that the marginal models (1) and (2) imply restrictions on the range of possible probabilities in the conditional models $P(A_1|A_0, X)$. Given assumptions 1-5 and marginal models (1) and (2), the joint distribution $P(A_0 = 1, A_1 = 1|X = x)$ is bounded by $\psi_0(x)\psi_1(x) \leq P(A_0 = 1, A_1 = 1|X = x) \leq \min\{\psi_0(x), \psi_1(x)\}$. This implies the following bounds on the conditional distribution

$$\psi_1(x) \leq P(A_1 = 1|A_0 = 1, X = x) \leq \min\left\{1, \frac{\psi_1(x)}{\psi_0(x)}\right\}.$$

Denote the upper bound by $\Delta_U(x) = \min\left\{1, \frac{\psi_1(x)}{\psi_0(x)}\right\}$. We propose the following model:

$$P(A_1 = 1|A_0 = 1, X = x) = \psi_1(x) + \phi\{\Delta_U(x) - \psi_1(x)\},\ 0 \leq \phi \leq 1. \tag{3}$$

**Interpretation of $\phi$.** If $\phi = 0$, then $P(A_1 = 1|A_0 = 1, X = x) = \psi_1(x)$, for all $x$, which is the conditional independence assumption (i.e. $A_1$ independent from $A_0$ given $X$). If $\phi = 1$, then $P(A_1 = 1|A_0 = 1, x) = \Delta_U(x)$, which is the largest possible probability that is compatible with the marginal distributions. For example, if more people with covariates $x$ would comply with the new treatment than the standard treatment (i.e., $\psi_1(x) > \psi_0(x)$), then setting $\phi = 1$ would be making the assumption that everyone with covariates $x$ that would comply with the standard treatment would also comply with the new treatment. If compliance with the new treatment is less likely than compliance with the standard treatment, among those that have covariates $x$, then setting $\phi = 1$ implies that $\psi_1(x)/\psi_0(x) \times 100$ percent of the subjects that would comply with the standard treatment would also comply with the new treatment. As an extreme example, if $\psi_1(x) = 0$, then no one with covariates $x$ would comply with the new treatment, and therefore $P(A_1 = 1|A_0 = 1, X = x) = 0$.

The marginal distributions (1) and (2) and conditional probability (3) imply a model for $P(A_1 = 1|A_0 = 0, X = x)$. Specifically, the probability $P(A_1 = 1|X = x)$ can be written as

$$P(A_1 = 1|X = x) = \sum_{a=0}^{1} P(A_1 = 1|A_0 = a, X = x)P(A_0 = a|X = x),$$

which implies

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

Solving the above equation for $P(A_1 = 1|A_0 = 0, X = x)$ results in

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

Our assumed association model is therefore

$$\begin{aligned} P(A_1 = 1|A_0 = a, X = x) &= a[\psi_1(x) + \phi\{\Delta_U(x) - \psi_1(x)\}] \\ &+ (1 - a)\frac{\psi_1(x) - [\psi_1(x) + \phi\{\Delta_U(x) - \psi_1(x)\}]\psi_0(x)}{1 - \psi_0(x)} \end{aligned} \tag{4}$$

This specification ensures that conditions 1 – 3 are met.

# 5 Inference methods

Inferences are based on an observed-data posterior distribution. We begin by writing a full-data likelihood for the joint distribution of all potential outcomes and covariates, $[Y_0, Y_1, A_0, A_1, X, Z]$. Priors are placed on the full-data parameters. The observed-data posterior is obtained by integrating the full-data model over the distribution of missing outcomes.

## 5.1 Full data likelihood

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 $\theta$ 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 (where $f(\cdot \mid \cdot)$ denotes a model for conditional distribution),

$$f(Y_0, Y_1, A_0, A_1, Z, X \mid \theta, \beta, \xi) = f(Y_0, Y_1 \mid S, Z, X, \theta) \times f(A_0, A_1 \mid Z, X, \beta) \times f(X, Z \mid \xi). \tag{5}$$

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 $(\theta, \beta, \xi)$ to the joint distribution (5), evaluated at $F_i = (Y_{0i}, Y_{1i}, A_{0i}, A_{1i}, Z_i, X_i)$. The full-data likelihood 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). Before proceeding, we make two assumptions about the full-data likelihood. First, we assume the potential outcomes are jointly independent of $Z$ and $X$ within principal stratum, so that

$$f(Y_0, Y_1 \mid S, Z, X, \theta) = f(Y_0, Y_1 \mid S, \theta). \tag{6}$$

This implies 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, we assume the potential outcomes related to compliance are jointly independent of $Z$, conditionally on $X$,

$$f(A_0, A_1 \mid Z, X, \beta) = f(A_0, A_1 \mid X, \beta). \tag{7}$$

Because $A_0$ and $A_1$ jointly determine $S$, this assumption means that conditionally on $X$, membership in principal stratum $S$ is independent of treatment assignment $Z$ (randomization assumption).

To parameterize the joint distribution of potential responses in (6), observe that there are four 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 distribution is represented on Table 1. Notice that $\theta_{10}(0) = \theta_{01}(0) = 0$ because under no exposure to any intervention, the potential outcomes cannot differ (exclusion restriction). Hence

$$\begin{aligned} f(Y_0, Y_1 \mid S, \theta) &= \left\{ \theta_{00}(0)^{(1-Y_1)(1-Y_0)}\ \theta_{11}(0)^{Y_1 Y_0} \right\}^{I(S=0)} \\ &\quad \times \prod_{s=1}^{3} \left\{ \theta_{00}(s)^{(1-Y_1)(1-Y_0)}\ \theta_{01}(s)^{(1-Y_0)Y_1}\ \theta_{10}(s)^{Y_0(1-Y_1)}\ \theta_{11}(s)^{Y_0 Y_1} \right\}^{I(S=s)}. \end{aligned}$$

The model for potential compliance variables given in (7) 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, from (1), the first factor is the Bernoulli mass

Table 1: Parameters indexing the joint distribution of potential outcomes within each principal stratum, $[Y_0, Y_1 \mid S]$. Each row of probabilities sums to one.

| | | | $(Y_0, Y_1)$ | | | |
|---|---|---|---|---|---|---|
| $A_0$ | $A_1$ | $S$ | (0,0) | (0,1) | (1,0) | (1,1) |
| 0 | 0 | 0 | $\theta_{00}(0)$ | 0 | 0 | $\theta_{11}(0)$ |
| 1 | 0 | 1 | $\theta_{00}(1)$ | $\theta_{01}(1)$ | $\theta_{10}(1)$ | $\theta_{11}(1)$ |
| 0 | 1 | 2 | $\theta_{00}(2)$ | $\theta_{01}(2)$ | $\theta_{10}(2)$ | $\theta_{11}(2)$ |
| 1 | 1 | 3 | $\theta_{00}(3)$ | $\theta_{01}(3)$ | $\theta_{10}(3)$ | $\theta_{11}(3)$ |

function $f(A_0 \mid X) = \psi_0(X)^{A_0}\{1-\psi_0(X)\}^{1-A_0}$. The second factor also is Bernoulli with probability given by (3).

To round out the model specification, we assume the parameters $(\theta, \beta)$ are a-priori jointly independent of $\xi$; i.e., $p(\theta, \beta, \xi) = p(\theta, \beta)\, p(\xi)$. A consequence of this is that the full-data posterior factors over $(\theta, \beta)$ and $\xi$,

$$\begin{aligned} p(\theta, \beta, \xi \mid F) &\propto \left\{ p(\theta, \beta) \prod_{i=1}^{n} f(Y_{0i}, Y_{1i} \mid S_i, \theta)\; f(A_{0i} \mid X_i, \beta) f(A_{1i} \mid A_{0i}, X_i, \beta) \right\} \\ &\quad \times \left\{ p(\xi) \prod_{i=1}^{n} f(X_i, Z_i \mid \xi) \right\} \\ &= p(\theta, \beta \mid F)\; p(\xi \mid F). \end{aligned}$$

We will see momentarily that the *observed*-data posterior also factors over $(\beta, \theta)$ and $\xi$, obviating the need to specify $f(X, Z \mid \xi)$ and $p(\xi)$.

### 5.2 Observed-data likelihood and posterior

The observed data likelihood is obtained by integrating over the sample space of missing potential outcomes. For those randomized to treatment group $Z$, the pair $(Y_Z, A_Z)$ is observed and $(Y_{1-Z}, A_{1-Z})$ is missing. Equivalently, we can say that $\{Y_Z, \mathcal{S}(Z, A_Z)\}$ is observed, where $\mathcal{S}(Z, A_Z)$ is the subset of principal strata compatible with the value of $A_Z$. For example, if $Z = 0$ and $A_Z = A_0 = 0$, then $(A_0, A_1) \in \{(0,0), (0,1)\}$, or equivalently $S \in \{0, 2\}$ (see Table 1). Hence $\mathcal{S}(0,0) = \{0, 2\}$.

For an individual on which $(Y_Z, A_Z, X)$ is observed, the contribution to the observed-data

likelihood is proportional in the model parameters to

$$\begin{aligned} f(Y_Z, A_Z, Z, X \mid \theta, \beta, \xi) &= f(X, Z \mid \xi) \iint f(Y_0, Y_1 \mid A_0, A_1, \theta)\, f(A_0, A_1 \mid X, \beta)\, dY_{1-Z}\, dA_{1-Z} \\ &= f(X, Z \mid \xi) \sum_{s \in \mathcal{S}(Z, A_Z)} \sum_{y=0}^{1} f(Y_Z, Y_{1-Z} = y \mid S = s, \theta)\, P(S = s \mid X, \beta) \end{aligned} \tag{8}$$

Hence the observed-data likelihood contribution for inference about $(\theta, \beta)$ is simply the double sum in (8), denoted from here by $L_{\text{obs}}(\theta, \beta \mid Y_Z, A_Z, Z, X)$. If $p(\theta, \beta, \xi) = p(\theta, \beta)p(\xi)$, then the observed-data posterior also factors over $(\theta, \beta)$ and $\xi$, and neither $f(X, Z \mid \xi)$ nor $p(\xi)$ needs to be specified to draw inference about the causal parameters $\theta$.

The observed-data likelihood can be simplified and written in terms of the observed outcome and compliance variables $Y$ and $A$ by considering each $(Z, A)$ combination. For example, if $Z = 0$ and $A = 0$, then $A_0 = 0$, $\mathcal{S}(0, 0) = \{0, 2\}$, and

$$\begin{aligned} L_{\text{obs}}(\theta, \beta \mid Y, A = 0, Z = 0, X) &= \sum_{s \in \{0,2\}} \sum_{y=0}^{1} f(Y_0, Y_1 = y \mid S = s)\, P(S = s \mid X) \\ &= \sum_{s \in \{0,2\}} \left\{ \theta_{00}(s)^{1-Y} \theta_{10}(s)^{Y} + \theta_{01}(s)^{1-Y} \theta_{11}(s)^{Y} \right\} P(S = s \mid X). \end{aligned}$$

A reparameterization leads to further simplification. For example, notice that $\pi_0(s) = \theta_{10}(s) + \theta_{11}(s)$ and $1 - \pi_0(s) = \theta_{00}(s) + \theta_{01}(s)$, for $s \in \{0, 2\}$. Then

$$L_{\text{obs}}(\pi, \beta \mid Y, A = 0, Z = 0, X) = \sum_{s \in \{0,2\}} \left[ \pi_0(s)^{Y} \{1 - \pi_0(s)\}^{(1-Y)} \right] P(S = s \mid X).$$

If we take a similar approach for the other $(Z, A)$ combinations, it is straightforward to show that

$$L(\pi, \beta \mid Y, A, Z, X) = \sum_{s=0}^{3} \pi_Z(s)^{Y} \{1 - \pi_Z(s)\}^{1-Y}\, P(S = s \mid X, \beta)\, G(s, A, Z),$$

where

$$\begin{aligned} G(s, A, Z) &= I(s = 0)(1 - A) + I(s = 1)\{A(1 - Z) + (1 - A)Z\} \\ &\quad + I(s = 2)\{AZ + (1 - A)(1 - Z)\} + I(s = 3)AZ. \end{aligned}$$

Recall that $P(S = s \mid X, \beta)$ is parameterized in terms of $\lambda$ and $\phi$. The observed data likelihood is therefore a function $(\lambda_0, \lambda_1, \phi)$ and the 7 parameters captured by $\pi$ (note that $\pi_0(0) = \pi_1(0)$ by exclusion restriction).

In our simulation and data analysis, we used WinBUGS to obtain draws from the observed-data posterior. Specifics about priors are given in Sections 6 and 7.

## 6 Simulation study

We carried out a simulation study to assess the performance of the proposed model in finite samples, and to quantify the benefit of using compliance-predictive covariates. We first simulated group membership $Z_i$ from a Bernoulli distribution with probability 0.5. We generated one covariate: $X_{1i} \sim N(0, 1)$ for $i = 1, \cdots, n$. The compliance indicators were generated as follows. First, $A_0$ for subject $i$ was generated from a Bernoulli with probability $\text{logit}^{-1}(0.4 + cX_{1i}/4)$. The marginal distribution of $A_1$ was Bernoulli with probability $\text{logit}^{-1}(0.3 + cX_{1i}/2)$; we considered $c = 1$ and $c = 2$. The model with $c = 2$ implies a stronger relationship between the covariate in compliance in both arms. With either model, overall compliance was about 60% in each arm on average. The variable $A_1$ was generated from the conditional distribution of $[A_1|A_0, X_1; \phi]$, given in (4). The observed response variable, given $S = s$ and $Z = z$, was set equal to 1 with probability $\pi_z(s)$, where $\pi_0(0) = \pi_1(0) = 0.1$, $\pi_0(1) = 0.4$, $\pi_1(1) = 0.2$, $\pi_0(2) = 0.2$, $\pi_1(2) = 0.4$, $\pi_0(3) = 0.4$ and $\pi_1(3) = 0.6$. We used $n = 250$ and generated data for $\phi = 0.1$ and $\phi = 0.9$.

We compared three different model specifications. For each fitted model, we used a burn-in of 500, and ran an additional 2000 draws, on 2 parallel chains. The first used the class membership $S$ as if it were observed (Model 1). In practice $S$ cannot be observed; however Model 1 is a useful reference for evaluating the performance of the proposed models. We assumed $Y$ given $S$ and $Z$ was Bernoulli with probability $\pi_z(s)$, for $z = 0, 1$, $s = 0, 1, 2, 3$ and $\pi_0(0) = \pi_1(0)$, and used Uniform $(0, 1)$ priors for the $\pi$'s.

Next, we fitted two different models to the observed data, using the model given in Section 4 and inference methods described in Section 5. The first of these two did not use the covariate $(X_1)$

in the compliance models (Model 2). That is, $P(A_0 = 1) = \psi_0$ and $P(A_1 = 1) = \psi_1$. The values of $\lambda$ and $\phi$ that were used in generating the data imply a corresponding value of the association parameter, denoted by $\phi^*$, in the model without covariates; it satisfied the identity

$$\psi_1 + \phi^*(\Delta_U - \psi_1) = \int [\psi_1(x) + \phi\{\Delta_U(x) - \psi_1(x)\}]\, dF(x),$$

where $\psi_z = \int \psi_z(x) dF(x)$, $z = 0, 1$, and $\Delta_U$ =min$(1, \psi_1/\psi_0)$. The identity can easily be solved for $\phi^*$ by approximating the integrals using a Gaussian quadrature. We fitted the model to the data using the $\phi^*$ that was implied by each set of $\lambda$ and $\phi$, so that each model could be viewed as correctly specified.

The next model we fitted included $X_1$ as a predictor of compliance, namely $P(A_0 = 1) =$ logit$^{-1}(\lambda_{00} + \lambda_{01}X_1)$ and $P(A_1 = 1) =$ logit$^{-1}(\lambda_{10} + \lambda_{11}X_1)$ (Model 3). In each case, we assumed the conditional distribution of $[A_1|A_0]$ followed (4), and that the true value of $\phi$ was known. In each model, relatively flat priors were used for each $\lambda$ (normal with mean 0 and variance 1000).

Under each scenario, 1000 data sets were generated and analyzed. The data were generated in R and draws from the posterior were obtained by calling OpenBUGS from R using the package BRugs. We then examined the frequentist properties of these Bayesian estimators of $\pi_1(3) - \pi_0(3)$, causal effect of the experimental treatment versus the active control among those who would comply with either. The posterior medians, width of the 90% credible intervals and whether or not the interval included the true value were recorded. These were summarized in Table 2 as the average posterior median across the 1000 samples (median), the standard deviation of these posterior means (ESD), the proportion of intervals that included the true value of the parameter (coverage) and the average width of the intervals (width).

**Results.** For $\phi = 0.9$, models 2 and 3 had a fairly similar performance. The posterior median of $\pi_1(3) - \pi_0(3)$ was close to its true value, and ESDs were about 0.09. The credible intervals were a little wider in the model without covariates, particularly when covariates were strongly predictive of compliance ($c = 2$). For both models 2 and 3, the credible interval coverage rates were greater than their nominal levels (93-97 percent coverage instead of 90 percent coverage), but were closer

to the nominal level for model 3. This suggests that for values of $\phi$ close to 1, including covariates in the model yields relatively modest benefits. For $\phi$ equal to 0.1, we found a stronger benefit from including covariates. Models 2 and 3 both produced posterior medians that were slightly below the true value of the parameter, on average. Both models were conservative in terms of coverage probability, but the model that included covariates had better coverage (closer to nominal level). The width of the interval was substantially smaller for the model that included covariates, especially when the relationship between the covariate and compliance was strong (a 32% reduction when $c = 2$). In general, we see that when $\phi$ is near 0, including a covariate in the model that is predictive of compliance reduces the width of the credible intervals. It is worth bearing in mind that $\phi$ captures association between $A_0$ and $A_1$ *conditional on covariates that are prognostic of compliance;* therefore in many applications we can expect that $\phi$ is bounded away from 1.

# 7 Analysis of commit to quit data

## 7.1 Data

The CTQ study was described in Section 2. The outcome was defined as continuous abstinence during the course of the trial, a common endpoint 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 $S = 3$ could be thought of as those who would be highly compliant with either intervention during the pre-quit week phase of the trial.

## 7.2 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, number of cigarettes smoked per day, age started smoking), previous quit attempts (number of quit attempts, ever quit for 24 hours), need for and concern about weight loss (body mass index, percent fat, frequency of

dieting, weight fluctuation in a typical week). We selected a compliance model for each arm based on parsimony, predictive accuracy and model fit, using the usual model selection techniques for logistic regression models. For compliance with the contact condition, we decided on the following model for $\psi_0(x)$,

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

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, `ed12` is the indicator that the subject has at least 12 years of education, and `ed15` is the indicator that the subject had at least 15 years of education (i.e., some college). This model appeared to fit 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

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

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, $z = 0, 1$, $s = 0, \cdots, 3$. Proper, but relatively flat normal priors were assumed for the $\lambda$'s, with mean 0 and variance 1000. At each fixed value of $\phi$, 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 1,000th draw, based on trace plots and the Gelman-Rubin statistic.

### 7.3 Results

The value of $\phi$ affects the estimated number of subjects in each stratum (i.e., the size of each compliance-type subpopulation). In Figure 2 we display estimates of the proportion of subjects in each principal stratum at various values of $\phi$. 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 $\phi$ was near

0, there were a non-trivial number of subjects who would comply in one arm but not the other (strata 1 and 2). When $\phi$ was near 1, nearly everyone that would comply in one arm would also comply in the other. Therefore, the estimated proportion of subjects in strata 0 and 3 increased as $\phi$ increased.

In Table 3 we present the 2.5, 50 and 97.5 percentiles of the posterior distribution of $\lambda_0$ and $\lambda_1$, under the assumption of conditional independence ($\phi = 0$); the results were similar at other values of $\phi$. For the contact compliance model, older subjects and those with more education appeared to be 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 were subjects who had fewer years of education.

The 2.5, 50 and 97.5 percentiles of the posterior distribution of the $\pi_z(s)$ parameters are given in Table 4. Among subjects that would be highly compliant with either intervention, we estimated that the abstinence rate in the exercise arm was about 0.33, compared to about 0.16 for the contact condition arm. The width of the corresponding credible intervals depends on the value of $\phi$; the intervals were the narrowest when $\phi$ was near 1, due to the fact that membership in stratum $S = 3$ would be largest when $\phi = 1$ (see Figure 2). Overall quit rates appeared to vary from about 4% (for those in stratum $S = 0$) to about 34% (for those in $S = 2$ or 3 who were randomized to receive exercise).

**Causal effect of exercise among compliers**. Primary interest is in $\pi_1(3) - \pi_0(3)$, the difference in cessation rates between exercise and contact groups for the subgroup that would comply with either intervention ($S = 3$). The posterior median, along with 95% credible interval for this causal risk difference is displayed in Figure 3 (b), at various values of $\phi$. For all values of $\phi$ the posterior median is positive, indicating an estimated higher probability of cessation if randomized to the exercise condition, among highly compliant subjects. For values of $\phi$ near 0, the 95% interval is very wide, indicating that we do not have much information about the intervention

effect. A value of $\phi$ between 0.5 and 1 might be more plausible, however, as the two interventions are similar in terms of what is expected from participants (both require attending a class 3 days per week). For large values of $\phi$, the interval becomes much more narrow, reflecting the fact that class membership in group 3 increases (see Figure 2). The interval overlaps the null value of 0, even when $\phi = 1$, but most of the values in the interval are greater than 0. The posterior median is 0.16 with 95% interval $(-0.04, 0.36)$ at $\phi = 1$.

Another way to compare treatments is to estimate the probability that cessation is higher if randomized to exercise, among highly compliant subjects ($S = 3$). We recorded $I\{\pi_1(s) > \pi_0(s)\}$ for each draw from the posterior. Here we focus on the $s = 3$ case. The average of these indicators over the 50,000 samples is an estimate of the probability of a benefit of exercise over wellness among compliers ($S = 3$). The results are presented in Figure 4 at various values of $\phi$. As $\phi$ increases towards 1, the benefit of exercise is more pronounced in the $S = 3$ stratum. If $\phi$ was near 0, the evidence is weaker, but the probability of a benefit from exercise was still about 0.78.

**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 that would comply with exercise, but not wellness. For the CTQ study, this population is of interest because anecdotally, many participants were attracted to the trial because of the exercise component. Figure 3 (a) displays posterior quantiles of $\pi_1(2) - \pi_0(2)$ as a function of $\phi$. As shown in Figure 1, when $\phi$ was near 0, group membership in $S = 2$ was at its largest. Therefore, as $\phi$ increased, the credible intervals for the ITT effect in arm $S = 2$ became wider. The majority of the posterior draws of the causal risk difference were greater than 0, suggesting a benefit of exercise for this subpopulation. However, the 95% credible intervals always included 0, indicating that we cannot rule out no exercise benefit. From Figure 4 we see that as $\phi$ decreased towards 0, it became highly likely that exercise was beneficial in $S = 2$ stratum. If $\phi$ is near 1, the evidence was weaker, but the probability of a benefit from exercise was still above 0.8.

We can also estimate the effect of being randomized to the exercise arm for strata the are

unions of the basic principal stratification. For example, investigators might 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., $S = 2$ or 3). In Figure 3 (c), we display the posterior median and 95% credible interval for the intention-to-treat effect in the $S = 2$ or 3 subpopulation, at various values of $\phi$. The interval did not overlap with 0 at any value of $\phi$, indicating a benefit of exercise for this subpopulation. Also, notice that the interval was much less sensitive to values of $\phi$ than were the effects for the population of compliers $S = 3$ or exercise-only compliers $S = 2$. The reason that the width of the 95% interval for $\pi_1(3) - \pi_0(3)$ changed dramatically as a function of $\phi$, was that the estimated size of the population of $S = 3$ strictly increased as $\phi$ increased (Figure 2). However, if one were to estimate the proportion of subjects in $S = 2$ or 3 from Figure 2, we would find that the number was nearly constant as a function of $\phi$. As $\phi$ increased, there was a shift from $S = 2$ to $S = 3$, but the combined population remained about the same. Because $S = 2$ or 3 included everyone that would comply with exercise, the information about that class comes from the marginal distribution of $A_1$.

**ITT effect for entire population**. The traditional intention-to-treat analysis can be thought of as the causal effect of $Z$ on the entire population (i.e., those in strata $S = 0, 1, 2$ or 3). The sensitivity parameter $\phi$ has no impact here, as varying $\phi$ does not shift the population $S \in \{0, 1, 2, 3\}$. The posterior median and 95% credible interval for the $\mathrm{E}(Y_1 - Y_0)$ was 0.085 $(0.002, 0.170)$, suggesting a benefit of being randomized to the exercise group instead of the wellness group.

**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 3. To illustrate, we estimated the average value of each covariate in each stratum when $\phi = 0.5$. We calculated the posterior probability that $S = s$ for every subject, at each value $s = 0, \cdots, 3$ and then obtained the sample mean of each covariate in stratum $s$ by weighting by the posterior probability that $S = s$. The results are shown in Table 5. Based on the results in the table, it appears that subjects that would comply with

the exercise condition but not the wellness condition ($S = 2$) were younger, more likely to have more than 15 years of education, were more likely to be employed and less likely to be married, than were subjects that would comply with the wellness condition, but not exercise ($S = 1$). The stratum consisting of subjects that would comply with either intervention ($S = 3$) had the highest percentage of subjects with more than 15 years of education.

### 7.4 The impact of covariates

We also explored the impact modeling compliance as a function of covariates by fitting a model with no covariates in compliance models (1) and (2). As described previously, without covariates the strata probabilities and causal effects are not point identified in this setting (2 active treatments). There are bounds, however, on these parameters. Figure 5 includes plots of the estimated posterior densities of $\pi_0(3)$, $\pi_1(3)$ and $\pi_1(3) - \pi_0(3)$ in the models with and without covariates, for $\phi$ 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 density. This reflected the fact that the data do not contain enough information 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 maximums. For $\phi$ equal to 1, as expected, the impact of covariates was less pronounced because at $\phi = 1$, a subject who complied with one intervention would also have complied with the other. Observed compliance is essentially all that is needed to identify the strata in that case. But since $\phi$ was unknown, Figure 5 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 $\phi$.

## 8 Discussion

Experimental studies with two active treatments are fairly common, especially in behavioral intervention studies. When there is non-compliance in each arm, statistical comparisons between the two interventions is challenging because the population of subjects that would comply with one

intervention might differ from the population that would comply with the other. We have proposed a model that utilizes 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 two marginal compliance models, conditional on covariates, and an association model that ensures 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. As demonstrated in the simulation study and in the data analysis, the inclusion of covariates in the compliance models effectively improves the precision of the estimates of the causal effects. In the smoking cessation analysis, we found that subjects who would comply with the exercise regimen would also tend to benefit from it. The evidence suggested that the exercise regimen would yield better results than the wellness therapy, among subjects that would comply with either intervention, although the results were not conclusive.

A philosophically satisfactory aspect of the principal stratification approach is that causal effects are always defined in terms of comparisons between counterfactual outcomes indexed by a variable over which the investigator has control. Using terminology from Pearl (2000), we can take an action "assign the subject to treatment A," but for some subjects, we cannot take the action "assign the subject to treatment A and have them comply with treatment A." Therefore, causal comparisons between active treatments for the entire population are not as well defined as causal comparisons between assigned treatments among meaningful subpopulations.

Our findings suggest that researchers who are planning trials with two 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 two treatments, beyond what could be learned from the usual intention to treat approach.

Areas of future research include developing methods for longitudinal data, where compliance and/or the response are measured repeatedly over time. In the CTQ trial, compliance and quit status were measured weekly. Because our focus was on continual abstinence, we needed to define compliance during the pre-quit day period (in order for the exposure to preceed the outcome). However, if one were interested in weekly quit status as an outcome, the methods proposed here could be extended to include weekly compliance data in order to estimate the principal effects at each visit. Similarly, if compliance status was observed at multiple time points, but the outcome was just observed at the end of the study (e.g., week 12 quit status), the ideas proposed here could be extended to investigate the effect of different types of compliance patterns on the outcome. For example, for longitudinal trials with one active treatment, Lin, Ten Have and Elliott (2006) describe a model based on latent compliance classes that characterize different compliance patterns over time; a similar approach might be reasonable here.

## Acknowledgements

This research was support by National Institutes of Health grant R01-HL-79457.

## 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. Bounds on causal effects in three-arm trials with non-compliance. *JRSSB*, in revision

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.

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

Lin, J. Y., Ten Have, T. R. and Elliott, M. R. (2006). Longitudinal nested compliance class model in the presence of time-varying noncompliance.

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.

Pearl, J. (2000). Causality: Models, Reasoning and Inference. Cambridge University Press.

Robins, J. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freedoman and A. Bailey (Eds), *Health Services Research Methodology: A Focus on AIDS*. Washington, D.C.: 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.

Table 2: Results from simulation study for inference about $\pi_1(3) - \pi_0(3)$ (true value is 0.2). Results are reported for true values of $\phi$ equal to 0.1 and 0.9, and for the two compliance models ($c = 1$ and $c = 2$). 1000 simulated data sets for each scenario were used.

| Model | median | ESD | width | coverage |
|---|---|---|---|---|
| | | $\phi = 0.1, c = 1$ | | |
| 1 | 0.19 | 0.10 | 0.32 | 0.90 |
| 2 | 0.17 | 0.11 | 0.89 | 1.00 |
| 3 | 0.16 | 0.15 | 0.75 | 0.98 |
| | | $\phi = 0.1, c = 2$ | | |
| 1 | 0.19 | 0.09 | 0.32 | 0.91 |
| 2 | 0.16 | 0.11 | 0.92 | 1.00 |
| 3 | 0.15 | 0.14 | 0.63 | 0.96 |
| | | $\phi = 0.9, c = 1$ | | |
| 1 | 0.20 | 0.08 | 0.27 | 0.90 |
| 2 | 0.21 | 0.09 | 0.37 | 0.96 |
| 3 | 0.20 | 0.09 | 0.35 | 0.95 |
| | | $\phi = 0.9, c = 2$ | | |
| 1 | 0.20 | 0.08 | 0.28 | 0.90 |
| 2 | 0.21 | 0.09 | 0.43 | 0.97 |
| 3 | 0.20 | 0.09 | 0.34 | 0.93 |

Table 3: 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 $\leq$ 12 yrs | -0.67 | 0.19 | 1.04 |
| education $\leq$ 15 yrs | -1.74 | -0.81 | 0.07 |
| exercise compliance model | | | |
| Intercept | -0.21 | -0.49 | 1.21 |
| education $\leq$ 12 yrs | -1.22 | -0.41 | 0.39 |
| education $\leq$ 15 yrs | -1.82 | -0.89 | 0.01 |

Table 4: Posterior median and 95% credible interval for the cessation probability parameters at $\phi$ equal 0, 0.5 and 1.

| Parameter | $\phi = 0$ | $\phi = 0.5$ | $\phi = 1$ |
|---|---|---|---|
| $\pi_0(0)$ | 0.04 (0.002, 0.12) | 0.04 (0.002, 0.10) | 0.05 (0.02, 0.10) |
| $\pi_0(1)$ | 0.24 (0.03, 0.48) | 0.33 (0.04, 0.70) | 0.34 (0.04, 0.83) |
| $\pi_0(2)$ | 0.07 (0.004, 0.23) | 0.09 (0.004, 0.33) | 0.09 (0.004, 0.55) |
| $\pi_0(3)$ | 0.16 (0.01, 0.52) | 0.13 (0.01, 0.37) | 0.18 (0.07, 0.32) |
| $\pi_1(0)$ | 0.04 (0.002, 0.17) | 0.04 (0.002, 0.15) | 0.04 (0.002, 0.13) |
| $\pi_1(1)$ | 0.14 (0.02, 0.36) | 0.19 (0.02, 0.53) | 0.26 (0.03, 0.79) |
| $\pi_1(2)$ | 0.33 (0.06, 0.63) | 0.35 (0.03, 0.79) | 0.36 (0.03, 0.88) |
| $\pi_1(3)$ | 0.34 (0.04, 0.74) | 0.33 (0.08, 0.59) | 0.33 (0.18, 0.50) |

Table 5: Estimated average value of covariates in each principal stratum.

| Covariate | $S = 0$ | $S = 1$ | $S = 2$ | $S = 3$ |
|---|---|---|---|---|
| Age | 38.5 | 44.0 | 37.8 | 42.2 |
| Education $\leq 12$ yrs | 0.42 | 0.50 | 0.26 | 0.30 |
| Education $13 - 15$ yrs | 0.42 | 0.30 | 0.41 | 0.30 |
| Education $> 15$ yrs | 0.16 | 0.20 | 0.33 | 0.40 |
| Employed | 0.82 | 0.57 | 0.83 | 0.72 |
| Married | 0.50 | 0.63 | 0.44 | 0.48 |

Figure 2: Estimated proportion of population in each principal stratum at each value of $\phi$.

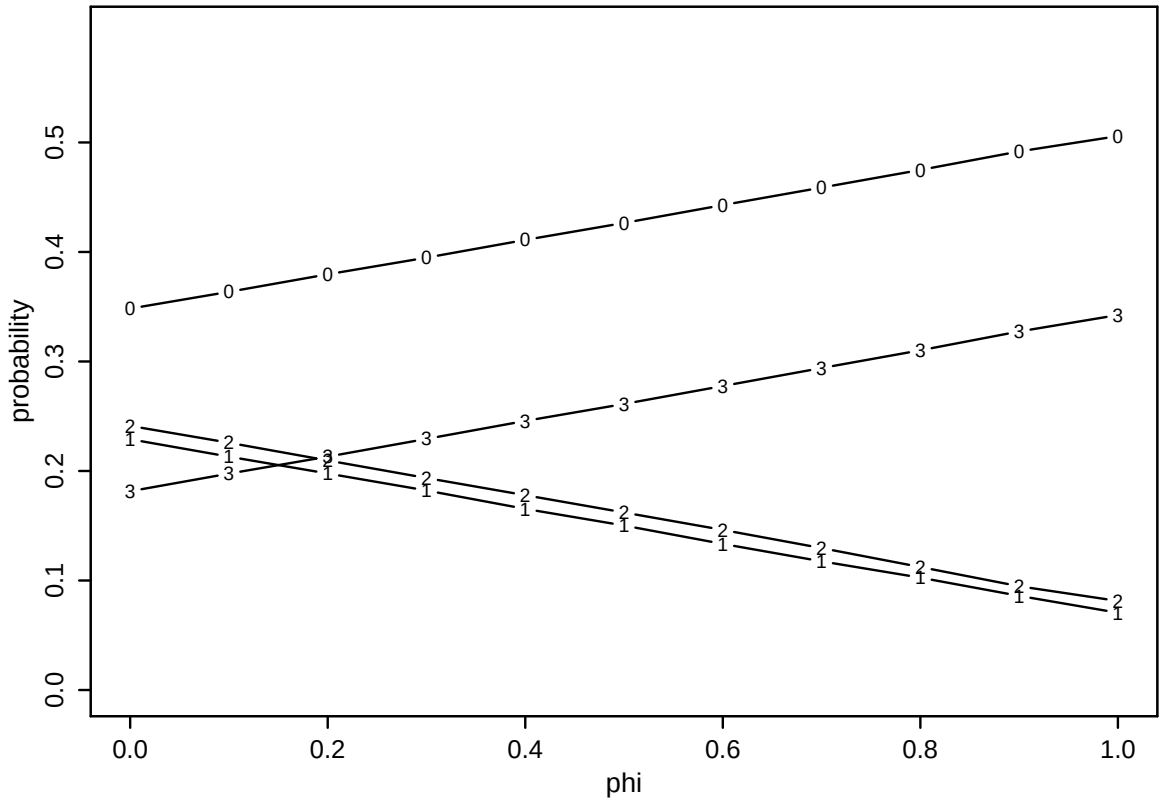


Figure 3: Plot of 2.5, 50 and 97.5 percentiles of posterior distribution of the risk difference for stratum 2, $\pi_1(2) - \pi_0(2)$, in figure (a), for stratum 3, $\pi_1(3) - \pi_0(3)$, in figure (b), and the combined stratum 2 and 3, $\pi_1(S \in \{2,3\}) - \pi_0(S \in \{2,3\})$,in figure (c), at different values of $\phi$.

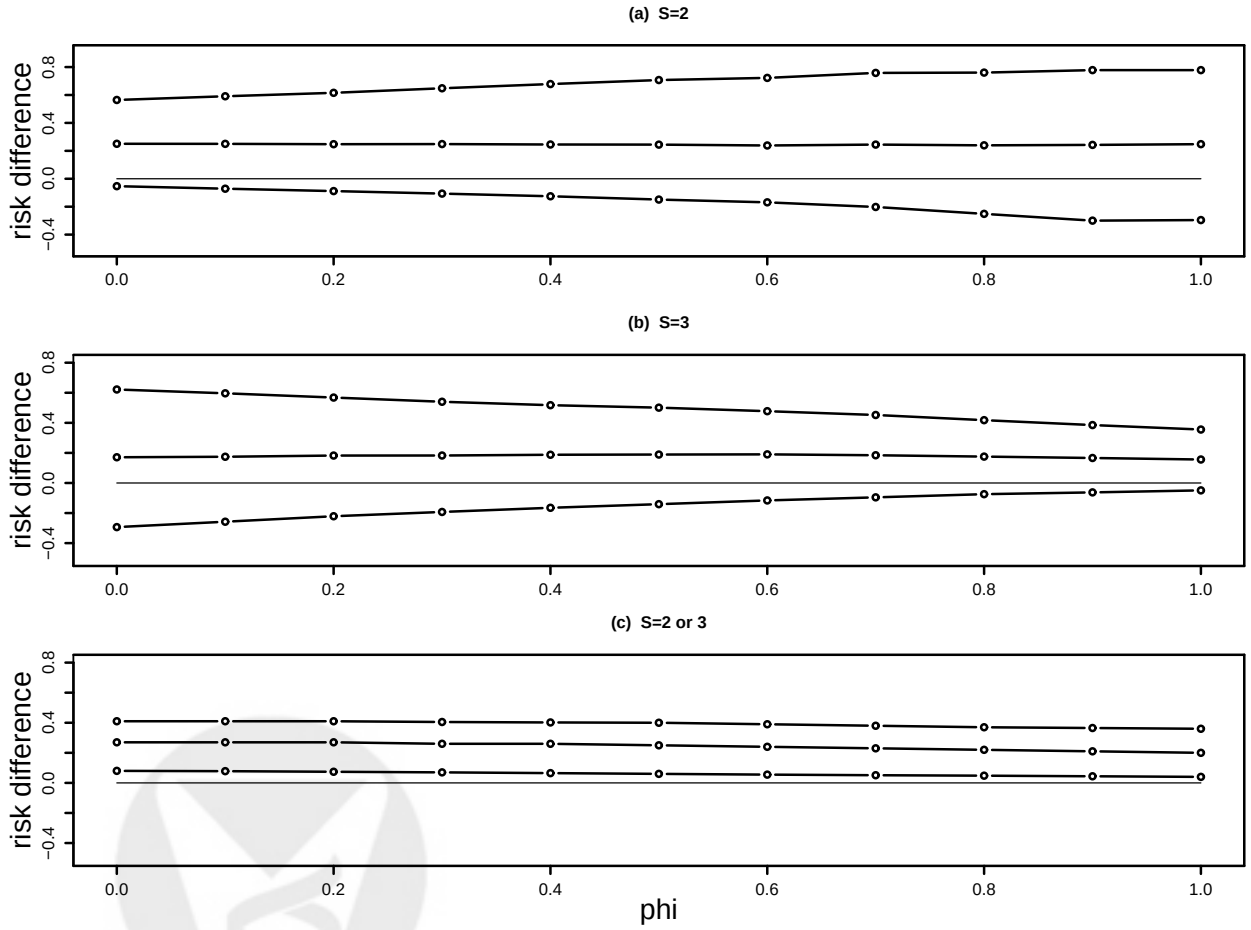


Figure 4: Plot of the posterior mean of $I(\pi_1(s) > \pi_0(s))$, for strata $s = 2, 3$, at different values of $\phi$. This all can be viewed as an estimate of the probability that cessation is more likely if assigned exercise in each stratum.

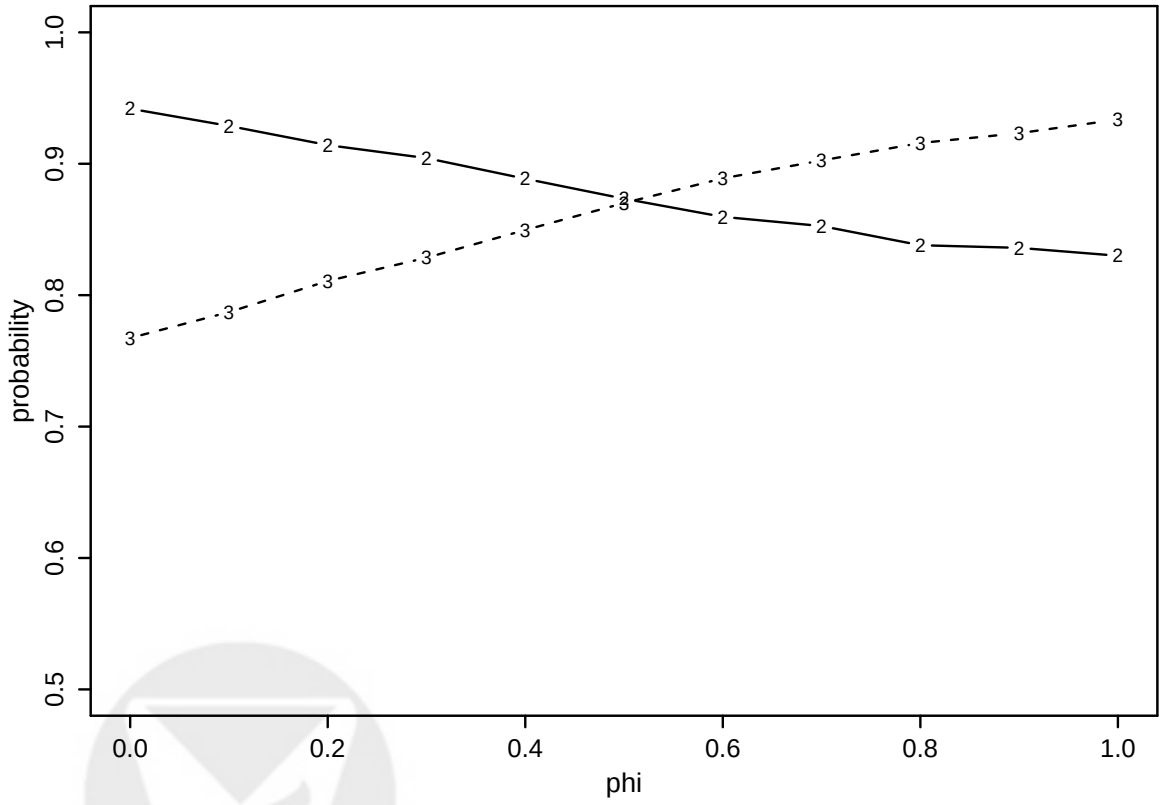


Figure 5: Plot of posterior distribution of $\pi_0(3)$, $\pi_1(3)$ and $\pi_1(3) - \pi_0(3)$ for the models with and without covariates, and at $\phi = 0$ and $\phi = 1$