Published in final edited form as:

Obs Stud. 2016 March ; 2: 29–50.

Patient Centered Hazard Ratio Estimation Using Principal Stratification Weights: Application to the NORCCAP Randomized Trial of Colorectal Cancer Screening

Todd A. MacKenzie[1], **Magnus Løberg**[2], and **A. James O'Malley**[1]

[1]Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, U.S.A

[2]University of Oslo, Oslo, Norway

Abstract

In randomized trials, the most commonly reported method of effect estimation is intention-to-treat (ITT), and to a lesser extent the per-protocol. The ITT is preferred because it is an unbiased estimator of the effect of treatment assignment. However, if there is any non-adherence the ITT is a biased estimate of the treatment effect, defined as the contrast between the potential outcome if treated versus the potential outcome if not treated. The treatment effect is most relevant to patients. Principal stratification is a framework for estimating treatment effects that combines potential outcomes and latent adherence strata. It yields an unbiased estimator of the complier average causal effect (CACE) for a difference in means or proportions, in the setting of all-or-nothing adherence. This paper addresses estimation of the causal hazard ratio for the compliers in a setting of right censoring of a time-to-event. We propose a novel approach to operationalizing principal stratification using weights. We report the results of simulations that vary the amount of adherence and selection bias that show the hazard ratio estimators we propose have minimal bias compared to the ITT, and per-protocol estimators. We demonstrate the approach using a population based randomized controlled trial of colorectal cancer screening subject to a high frequency of nonadherence in the screening arm.

Keywords

Time-to-event analysis; Cox model; Instrumental Variables; Semi-parametric

1. Introduction

Randomized clinical trials are the gold standard of comparative effectiveness studies. A highly regarded product of randomized studies is the intention-to-treat (ITT) estimator, which compares the endpoint between two or more levels of randomization. In a study in which the levels are treatment and control (no treatment), the ITT estimator is an unbiased estimate of the effect of treatment assignment, but biased as an estimate of treatment efficacy (unless there is perfect adherence). The effect of treatment assignment has smaller absolute value than the effect of treatment (assuming the effect of treatment assignment is strictly through its effect on treatment recieved). It will be smaller to the extent that there is lack of compliance. As an alternative to the ITT estimate it is common for clinical investigators to report either or both of the as-treated estimate and the per-protocol estimate. However, both

are subject to selection bias if subjects select their treatment as opposed to taking the randomly assigned treatment. Randomized trials without perfect adherence are, in part, observational studies Hernan et al. (2013).

The effect of treatment if subjects were to comply is commonly referred to as the *complier average causal effect* (CACE) or the *local average treatment effect* (LATE). It directly addresses the question; what is the difference in outcomes between taking versus not taking the treatment being considered? For this reason we refer to it as *patient centered* to distinguish it from the ITT approach which is system or policy centered. The definition of this effect requires a potential outcomes framework. The potential outcomes for a particular subject consist of one outcome for each of the possible assignment levels. The most common potential outcome framework is known as *principal stratification* (PS). The usual setting in which principal stratification is applied is the simple, and admittedly idealistic setting in which subjects are randomly assigned to treatment or not (control), and in which compliance is all or nothing. The principal strata are *compliers* (Co), those who will comply with their assigned treatment no matter what it is, *always takers* (AT), those who will take the treatment no matter what assigned to, *never takers* (NT), those who will not take the treatment regardless of assignment and *defiers*, those who will do the opposite of their assigned treatment). Generally it is assumed that there are no defiers, and furthermore this assumption is required in order to make the CACE identifiable. An excellent overview of all the estimators discussed above and their limitations is provided in Shrier et al. (2013).

A closely related area is instrumental variables (IVs). An instrument is a variable that effects treatment choice but has no effect on the outcome of interest except through its effect on the treatment chosen. Randomization is an example of an instrumental variable. In the setting of a linear causal model for the effect of treatment on the outcome, the estimator of the CACE derived using principal stratification is equivalent to the instrumental variable estimator which is also known as two stage least squares.

There has been relatively little work on IV or PS estimation methods for non-linear causal models. Using instrumental variables in the estimation of causal odds ratios has been addressed by Johnston et al. (2008) and Rassen et al. (2009). The estimation of the odds ratio in a principal stratification framework has been addressed by Cai et al. (2011). Baker (1998), Frangakis and Rubin (1999) and Nie et al. (2011) have derived estimators of survival in treated and untreated subjects using PS.

The hazard ratio is the favored method for comparing risk between two exposures for time-to-event outcomes. The estimation of the hazard ratio of Cox's proportional hazards model using an instrumental variable has been addressed by MacKenzie et al. (2014), and Wan et al. (2015). The estimate of hazard differences using instrumental variables, as in an additive hazards models has been developed by Li et al. (2015) and Tchetgen Tchetgen et al. (2015). For the setting of a RCT in which subjects randomized to control have no access to the experimental therapy, but subjects randomized to the experimental therapy may not take it at all, Loeys and Goetghebeur (2003) proposed an estimator of the hazard ratio for the effect of treatment. Cuzick et al. (2007) propose an estimator of the hazard ratio for the all-or-nothing setting in which those randomized to treatment may not adhere and those randomized to

control may receive the treatment (which they call contamination). In this paper we propose a new estimator of the causal hazard ratio and illustrate its ease of implementation using a novel weighting approach.

2. Methods

2.1 Notation

Let R represent the instrumental variable, which we assume to be binary. For example, $R = 0$ indicates randomization to the control, and $R = 1$ indicates randomization to treatment. Let the binary indicator or treatment received be $X = X(R)$ where $X(0)$ is the exposure if randomized to the control arm, and $X(1)$ is the exposure if randomized to the treatment arm. For example if there was perfect adherence, i.e. $X(0) = 0$ and $X(1) = 1$ for all subjects. Let Y represent an outcome of interest. Moreover, let $Y(1)$ and $Y(0)$ be the potential outcomes that would result were a subject to, respectively, receive, and not receive the treatment. The outcome may be multi-dimensional. For instance, the purpose of this paper is to cover the setting of right censored data in which the post-randomization data consists of both a time-to-event and censoring time. Let W be characteristics of the subjects at baseline.

2.2 Principal Stratification

Principal stratification starts with the fact there are 4 mappings from the set $\{0, 1\}$ to itself, as indicated in Table 1. Usually these mappings are given the suggestive names found in the final column of Table 1. For instance, *Compliers* (Co) are those subjects for which the value of the exposure is identical to the value of the instrument, *Never Takers* (NT) are those subjects for which $X = 0$ regardless of R, *Always Takers* are those subjects for which $X = 1$ regardless of R, and *Defiers* are those subjects for which $X = 1 - R$. Cuzick et al. (2007) refers to the always takers, compliers and never takers, as the insistors, ambivalents, and refusers, respectively. For a given experiment (e.g, study) subjects fall into one of the 4 strata. A key point is that these strata are not directly observeable, and therefore this stratification is latent. For instance, a subject for whom $R = X = 1$ appears to be a complier but could also from the strata AT.

The stratum of defiers, e.g., those who do the opposite of their assignment, is usually assumed to not exist. In other words, it is assumed that the probability of being a defier is zero (or close enough to zero to be irrelevant). This assumption is sometimes to referred to as *monotonicity* or *positivity*. By assuming no defiers it is now possible to partially identify the AT and NT strata. For instance, a subject for whom $R = 0$, $X = 1$ must be from the AT strata, and a subject for whom $R = 1$, $X = 0$ must be from the NT strata. Table 2 lists the possible strata a subject could be in according to their observed values for the instrument and exposure.

Let p_{Co}, p_{AT}, p_{NT} be the probabilities of sampling subjects from the three strata, Co, AT and NT, respectively. Under the assumption of no defiers these three probabilities sum to unity. The conditional distribution of X given R determines the strata probabilities: From Table 2 it follows that $p_{AT} = \Pr[X = 1|R = 0]$ and $p_{NT} = \Pr[X = 0|R = 1]$, and the equation, $\Pr[X = 1|R$

= 1] = p_{Co} + p_{AT} implies p_{Co} = Pr[X = 1|R = 1] – Pr[X = 1|R = 0] (or equivalently p_{Co} = Pr[X = 0|R = 0] – Pr[X = 0|R = 1]).

2.3 Assumptions

We assume that $\{Y(0), Y(1)\}$ and R are independent conditional on the principal strata,

$$\{Y(0), Y(1)\} \perp R \mid PS.$$

This assumption excludes the possibility that randomization has an effect on the potential outcomes in the always takers, never takers, or in the compliers. For example, it excludes the possibility of a placebo effect in the compliers. and it excludes the possibility that making it harder for always takers to get treatment affects $Y(1)$. This assumptions implies the following two properties which we utilize in a subsequent section

$$Y(1) \mid R = 0, PS = AT \sim Y(1) \mid R = 1, PS = AT \quad (1)$$

and

$$Y(0) \mid R = 0, PS = NT \sim Y(0) \mid R = 1, PS = NT. \quad (2)$$

2.4 Identifiability of Conditional Distributions Given Principal Strata

This section addresses identifiability, and estimation of the conditional distributions of $Y(0)$ and $Y(1)$ given the principal stratum. The essential point of this subsection is that subjects for whom $R = X = 0$ are a mixture of compliers and never takers, while subjects for whom $R = X = 1$ are a mixture of compliers and always takers. We can use distributions of observations on subjects for whom $R = X = 1$ to infer distributions on the treated compliers if we subtract out the always takers. Similarly, we can an estimate the distribution for non-treated compliers using subjects for which $R = X = 0$ if we subtract the distribution in the never takers.

The conditional distribution of $Y(1)$ given $PS = NT$ is not observable, as is the conditional distribution of $Y(0)$ given $PS = AT$. On the other hand,

$$\begin{aligned} &Y(1) \mid (PS = AT) \sim Y(X) \mid (R = 0, X = 1) \quad (3) \\ &= Y \mid (R = 0, X = 1) \end{aligned}$$

and

$$Y(0) \mid (PS = NT) \sim Y(X) \mid (R = 1, X = 0) \quad (4)$$
$$= Y \mid (R = 1, X = 0)$$

The conditional distribution of $Y(1)$ among those assigned to and exposed to treatment, that is, $Y(1)|R = 1, X = 1$, is a mixture of the distribution of $Y(1)$ among the compliers and its distribution among the always takers. In particular,

$$Y(1) \mid R = 1, X = 1 \sim \frac{p_{Co}}{p_{Co} + p_{AT}}[Y(1) \mid R = 1, PS = Co] + \frac{p_{AT}}{p_{Co} + p_{AT}}[Y(1) \mid R = 1, PS = AT].$$

This result combined with (1) implies

$$Y(1) \mid PS = Co \sim \frac{p_{Co} + p_{AT}}{p_{Co}}(Y(1) \mid R = 1, X = 1) - \frac{p_{AT}}{p_{Co}}(Y(1) \mid R = 0, X = 1) \quad (5)$$
$$= \frac{p_{Co} + p_{AT}}{p_{Co}}(Y \mid R = 1, X = 1) - \frac{p_{AT}}{p_{Co}}(Y \mid R = 0, X = 1).$$

In words, among the compliers the distribution of the outcome if a subject is treated is the distribution of the observed outcome among subjects for which $R = 1, X = 0$ weighted by $1 + \frac{p_{AT}}{p_{Co}}$ minus the distribution among subjects for which $R = 1, X = 0$ weighted by $\frac{p_{AT}}{p_{Co}}$.

Similarly, among the compliers the distribution of $Y(1)$, can be derived using (1) as the following weighted subtraction of distributions

$$Y(0) \mid PS = Co \sim \frac{p_{Co} + p_{NT}}{p_{Co}}(Y(0) \mid R = 0, X = 0) - \frac{p_{NT}}{p_{Co}}(Y(0) \mid R = 1, X = 0). \quad (6)$$

2.5 Estimating Distributions, and Their Parameterizations, on the Compliers

Formulas (5) or (6) can be used to estimate the (marginal) conditional distributions of the potential outcomes $Y(1)$ and $Y(0)$ in the compliers as follows: First, estimate the following conditional distribution of $Y|R = 1, X = 0$, $Y|R = 0, X = 0$, $Y|R = 0, X = 1$, and $Y|R = 1, X = 1$ using for instance the empirical distribution function (EDF). Second, substitute the estimates of each of these distributions into formulas (5) or (6). A similar approach could be used to estimate the distribution of baseline characteristics of the compliers (see Appendix A).

Suppose one is interested in a parameter that quantifies the difference between the distributions $Y(1)|PS = Co$ and $Y(0)|PS = Co$ in the compliers. For instance, suppose $Y(x)|PS = Co$ has a cumulative distribution function (CDF) given by $G(y; \theta = \beta x)$ for some

family of CDF's $\{G(y;\theta)\}_\theta$. Then one more step is needed to estimate the parameter β given estimates of the distributions $Y(0)|PS = Co$ and $Y(1)|PS = Co$ that were derived in the second step. To estimate β we can use any known mapping from those distributions to the parameter space of β. If the estimates of $Y(x)|PS = Co, x = 0, 1$, obtained in the second step are discrete, as they will be if EDFs are used in the first step, then weighted maximum likelihood or weighted least squares can be used to estimate β.

An heuristic argument for the consistency of estimators obtained using this approach is as follows; 1, the empirical distribution function of $Y|R = r, X = x$ based on the observed randomized trial data is a consistent estimator of the distribution (e.g. of the cumulative distribution function); 2, linear combinations of consistent estimators (of distributions) are consistent for the linear combinations of the corresponding underlying estimands (e.g. distributions), whether or not some of the weights are negative; 3, a method that maps empirical distribution functions to a parameter (e.g. the use of maximum likelihood or least squares) and is consistent will have the property that as any approximation to the underlying true distribution becomes closer to the true distribution the parameter mapped to becomes closer to the true parameter.

2.6 Principal Stratification Weighted Data Analysis

The key point of this section is that the principal stratification estimates of the distribution described in the previous section, is equivalent to the use of a certain weighted data set. In what follows we refer to these as *principal stratification weights*. Formula (5) states that the distribution of the random variable, $Y(1)|PS = Co$, can be obtained by mixing the distributions, $Y|R = 1, X = 1$ and $Y|R = 0, X = 1$ with weights $1 + \frac{p_{AT}}{p_{Co}}$ and $-\frac{p_{AT}}{p_{Co}}$ respectively. Similarly, formula (6) states that the distribution of $Y(0)|PS = Co$ can be obtained by mixing the distributions $Y|R = 0, X = 0$ and $Y|R = 1, X = 0$ with weights $1 + \frac{p_{NT}}{p_{Co}}$ and $-\frac{p_{NT}}{p_{Co}}$ respectively. There is a unique weight for each of the four strata $\{(R = r, X = x)\}_{r=0,1,x=0,1}$.

As the distribution of $Y(x)|R = r, X = x$ is estimable by the sample $\{Y_i = Y(X_i)\}_{R_i=r,X_i=x}$ it follows that the joint distribution of $[X, Y(X)]|PS = Co$ is estimable by the data $\{X_i, Y_i, W_i\}_{i=1}^n$ for which the weight

$$W_i = \begin{cases} \frac{c_1}{n_{1,1}}(1 + \frac{p_{AT}}{p_{Co}}) & \text{if } R_i = 1, X_i = 1 \\ -\frac{c_1}{n_{0,1}}\frac{p_{AT}}{p_{Co}} & \text{if } R_i = 0, X_i = 1 \\ \frac{c_0}{n_{0,0}}(1 + \frac{p_{NT}}{p_{Co}}) & \text{if } R_i = 0, X_i = 0 \\ -\frac{c_0}{n_{1,0}}\frac{p_{NT}}{p_{Co}} & \text{if } R_i = 1, X_i = 0 \end{cases}$$

where $n_{r,x}$ is number of subjects for which $R = r$, $X = x$, and c_0, c_1 are arbitrary positively valued constants, which equal the sum of the weights for subjects for whom $X = 0$ and $X = 1$ respectively. The estimators discussed below are invariant to the choice of c_0 and c_1. However to retain consistency with the convention that the sum of weights in a subgroup is the number of subjects in that subgroup (as it is when all weights equal 1) we take $c_0 = n_{\cdot 0}$ and $c_1 = n_{\cdot 1}$.

Our purpose for formulating the principal stratification approach in terms of weights is to estimate parametric or semi-parametric models for the treated and untreated compliers using standard estimation methods that facilitate weights. To estimate parameters of any model for the joint distribution $[X, Y(X)]|PS = Co$ we propose to use this weighted data set and apply conventional estimation methods.

Unlike other common weights such as frequency weights, or inverse probability weights some of the principal stratification weights are negative. In particular, the weights are negative for those subjects that are discordant with respect to the instrument and the exposure, $R \neq X$. Conceptually, the negative weighting accomplishes the following; it removes from the sample of subjects who appear to comply with assignment $R = X$ the always takers in the case of $R = X = 1$ and the never takers in the case of $R = X = 0$.

The estimates of the frequency of each principal strata are $\hat{p}_{AT} = \frac{n_{01}}{n_{00} + n_{01}}$, $\hat{p}_{NT} = \frac{n_{10}}{n_{10} + n_{11}}$ and $\hat{p}_{Co} = 1 - \hat{p}_{AT} - \hat{p}_{NT}$ which can be written $\{n_{00}n_{11} - n_{01}n_{10}\}/\{n_{0\cdot}n_{1\cdot}\}$. Using these estimates one can derive the expressions for the principal stratification weights (PSW) shown in Table 3. Note: Because of the choice $c_0 = n_{\cdot 0}$ and $c_1 = n_{\cdot 1}$ the diagonal cells are identical.

2.7 Right Censored Outcomes

The focus of this paper is inference for outcomes subject to right censoring. Let $\{T^0(x)\}_{x=0,1}$ be the potential time-to-events, if not treated, and treated, respectively. Let $\{C(x)\}_{x=0,1}$ be the corresponding potential censoring times, if treated and not treated. The data we observe is $\{R_i, X_i, \Delta_i = \Delta_i(X_i), T_i = T_i(X_i)\}$ where $\Delta_i(x) = 1[T_i^0(x) \leq C_i(x)]$ and $T_i(x) = \min\{T_i^0(x), C_i(x)\}$.

There are two distinct approaches we can take that depend on the order in which right censoring and principal stratification are accounted for. For instance, we could take into account right censoring as a first step, and estimate the conditional distribution of $T^0(x)|R = r, X = x$ for $r = 0, 1$, $x = 0, 1$ using for instance, Kaplan-Meier estimators, and next derive the conditional distributions $T^0(0)|PS = Co$ and $T^0(1)|PS = Co$ using formulas (5) and (6). The use of the Kaplan-Meier in the first step requires the assumption that conditional on $(R = r, X = x)$ the censoring time $C = C(X)$ is non-informative about the time-to-event $T^0 = T^0(X)$, e.g. $T^0(x)$ and $C(x)$ are independent given $(R = r, X = x)$ for $r = 0, 1$, $x = 0, 1$. This is the approach proposed in Baker (1998). For estimation of quantities other than the survival curves, such as the hazard ratio relating treated to untreated compliers, it is not obvious how

to proceed using these survival curves. We describe a method to operationalize this in Appendix B.

The approach we emphasize in this paper is to use the principal stratification first, and then conventional methods to account for right censoring. This approach is very easy to implement using PSWs. By employing the weights one is creating a sample from compliers. The steps are (i) calculate the weights and (ii) execute any statistical method for right censored data that allow weights. Possible tests and estimators that could be employed be are the log-rank, Peto-Peto-Prentice test, incidence estimators (see Appendix C), Kaplan-Meier estimator, or Cox's model to determine a hazard ratio as described in the next section. In each case the first step is the calculation of the PSW.

2.8 Assumptions of the Independence of Censoring

We assume that censoring is non-informative conditional on the principal strata. In particular, it is sufficient to assume that censoring is non-informative in the compliers.

$$\Pr\left[T(x) \geq s \mid T(x) \geq t, C(x) \geq t, \mathrm{PS} = \mathrm{Co}\right] = \Pr\left[\mathrm{T(x)} \geq \mathrm{s[T(x)} \geq \mathrm{t}, \mathrm{PS} = \mathrm{Co}\right] \quad (7)$$

for $x = 0, 1$, and $s \geq t$. It is this assumption that ensures consistency of the hazard ratio were one to randomize individuals from the complier strata. Given that the principal stratification weighting yields an estimate of the distribution of potential time-to-event and censoring in the compliers, it follows that this assumption is needed for consistency of our approach.

2.9 Patient Centered Hazard Ratio via Cox Regression with Principal Stratification Weights

In this section we present an estimator of the patient centered hazard ratio, which we define as the ratio of the hazard of a treated complier to the hazard of an untreated complier. In particular, we are estimating the parameter, β, in the following causal version of Cox's model,

$$\frac{\Pr\left[\mathrm{T(1)} \in [\mathrm{t}, \mathrm{t} + \mathrm{dt}] \mid \mathrm{T(1)} \geq \mathrm{t}, \mathrm{PS} = \mathrm{Co}\right]}{\Pr\left[\mathrm{T(0)} \in [\mathrm{t}, \mathrm{t} + \mathrm{dt}] \mid \mathrm{T(0)} \geq \mathrm{t}, \mathrm{PS} = \mathrm{Co}\right]} = \exp\left[\beta\right]. \quad (8)$$

where β is the log hazard ratio.

We propose to estimate the hazard ratio using weighted maximum partial likelihood estimation, in which the weights supplied are the principal stratification weights described in section 2.6. The weighted log-partial likelihood that is created using the weights yields an estimator of what the MPLE would be if the compliers were sampled from and assigned treatment was the instrument (randomization).

This weighted approach could make estimation of the hazard ratio very easily operationally, as most statistical software allows specification of weights. Unfortunately, the same software typically returns an error if it encounters a negatively valued weight. Therefore it may be

necessary for users to code software that allows weights less than zero. Appendix D contains the R code we used to calculate the hazard ratio using principal stratification weights.

2.10 Variance Calculation

The variance of the estimator of the causal hazard ratio we propose can be derived using counting process theory. In particular, the score equation for the maximum partial likelihood with weights can be used to derive a sandwich variance formula in which the numerator is obtained using the variation process of a counting process martingale, and the denominator is obtained as the second derivative of the maximum weighted partial likelihood.

3. Simulation

We performed simulations to evaluate the hazard ratio estimator we have proposed using principal stratification weights as described in the previous section, and to compare it to the conventional estimators, (i) intention-to-treat, (ii) as-treated and (iii) per-protocol. For each dataset simulated, the variables listed in Table 4 are randomly generated. The first variable to be generated, Z, is latent. Subsequently the principal strata are generated conditional on Z using a logistic model for a 3-level dependent variable in which the odds ratio is 5 and the cutoffs are determined by the specified proportions of always takers and never takers. The instrument (randomization) is generated from a Bernoulli(0.5). The exposure is determined based on the instrument and the principal strata. The true time-to-event is generated according to a Cox model with baseline hazard according to the prinicipal strata, and single covariate equal to the exposure received; $\Pr[T(x) = t|T(x) \geq t, PS] = \lambda_{PS}(t) \exp(\beta x)$. Censoring is generated according to a uniform distribution on the interval from zero to ρ; the parameter ρ is determined so as to achieve the specified frequency of censoring.

The random variables are generated based on the parameters specified in Table 5. For any one randomly generated dataset, the parameters are constant. The parameters varied between datasets as follows. The parameter representing the true hazard ratio took on 9 distinct values ranging from 1/5 to 5 with equal frequency. The number of events in a dataset was 50, 100, 150 or 200, with equal frequency. The parameter f_C represents the frequency of censoring and is sampled uniformly from the interval (0, 0.9). The frequencies of the always takers and never takers were sampled uniformly (and independently) from the interval between 0 and 0.4. The parameter, ϕ, relating the time-to-event to the latent variable determines the amount of confounding; if it is zero there is no counfounding; if $\phi < 0$ there is positive confounding (longer survival); if $\phi > 0$ there is negative confounding (shorter survival). The confounding is due to (i) the effect of the latent variable on the time-to-event and (ii) its association with the principal strata which causes an association with X.

3.1 Simulation Results

We classified hazard ratio estimates over 1000 or less than 1/1000 as being non-applicable. The frequency with which the PSW estimator was non-applicable ranged from 1% when the proportion of compliers was 90% to 12% when the complier proportion was as low as 20%.

Figure 1 depicts the bias of the estimators of the hazard ratio, derived using principal stratification weights, ITT, Per Protocol and As Treated estimators. It plots the geometric

mean over 5000 simulations of each of the hazard ratio estimators for each of the values of the true hazard ratio. An unbiased estimator would fall along the line of identity. The panel on the left shows the results when there is a combination of positive confounding and less than 75% compliance. As expected the Per Protocol and As Treated estimators are positively biased for the entire range of true hazard ratios. The ITT is biased toward the null. The PSW estimator is unbiased. The other PSW estimator explained in Appendix B is also represented in the graphic, and its results are essentially equivalent to the PSW estimator.

The panel on the right in Figure 1 plots the bias when there is negative confounding and compliance less than 75%. Whereas the PSW estimators are unbiased, the Per Protocol and As Treated are negatively biased, and the ITT is biased toward the null.

The simulations demonstrate that the hazard ratio estimators we proposed are unbiased for estimation of the effect of treatment in compliers, and in that regard vastly superior to the Per Protocol and As Treated estimators, as well as the ITT.

4. Example: Effect of Screening on Colorectal Cancer Incidence

Colorectal cancer is the third most common cancer in the world. The benefit of screening for colorectal cancer prevention is an active area of study. The Norwegian Colorectal Cancer Prevention Trial (NORCCAP) was a randomized population based trial of the effect of screening for colorectal cancer using flexible sigmoidoscopy Holme et al. (2014). The study included all men and women living in Oslo or Telemark County, Norway, a total of 100,210 individuals. Of these 20,780 were randomly selected to receive an invitation to undergo screening, and 79,430 individuals were assigned to be controls. Due to exclusions (individuals determined to have died, been diagnosed with colorectal cancer or emigrated before study entry) the numbers in each arm reduced to 20,572 and 78,220. Here we report data from the first 10 years of the study.

None of the individuals in the control arm had access to screening. In terms of principal strata this means that the proportion of Always Takers was zero, $\hat{p}_{AT} = 0$. Of the 20,572 who were invited to receive flexible sigmoidoscopy, 63.0% (N=12955) received screening, and 37.0% did not. Therefore the estimates of the proportions of Never Takers and Compliers are $\hat{p}_{NT} = 0.37$ and $\hat{p}_{Co} = 0.63$ respectively. The corresponding weights are given in Table 6. Since there are no Always Takers the subjects randomized to screening who undergo screening are all compliers, those who are randomized to screening but do not undergo screening are all Never Takers, and those who are randomized to no screening, and indeed do not undergo screening are a mix of Compliers and Never Takers.

The number of colorectal cancers diagnosed (and total years of follow-up) during the first 10 years were 889 (740555), 91 (69653) and 115 (125270), respectively among the assigned controls, screening arm participants who did not screen and screened participants respectively. The corresponding incidence rates were 1.2, 1.3 and 0.9 per 100 years. The later incidence is an estimate of incidence in a screened complier. We calculated the incidence in the unscreened compliers to be 1.2 using the weighting method described in Appendix C.

As reported in Holme et al. (2014) a significant reduction in colorectal cancer incidence was found in those invited to undergo screening. The intention-to-treat hazard ratio was 0.80 (95%CI:0.70–0.92). That is, over the 10 years of follow-up, individuals invited to screen were 20% less likely to be diagnosed with colorectal cancer. The corresponding reduction in those aged 50–54 was 32% (95%CI: 7% – 53%), and in those aged 55–64, was 14% (−2% −28%).

Holme et al. (2014) report a principal stratification based estimator; they report the CACE at 10 years, which they refer to as an estimator of the per-protocol effect (not to be confused with what is commonly referred to as the per-protocol estimator whose estimand has no convenient form) as well as the adherence adjusted analysis. They reported an adherence adjusted absolute risk difference of 0.42% (95% CI, 0.69% to 0.15%). In other words the number of compliers needed to treat to eliminate 1 colorectal cancer by 10 years is estimated to be 1.0/0.0042=238.

Figure 2 shows estimates of the proportion of subjects being diagnosed with colorectal cancer over the 10 years following study inception. The thick dashed line is the Kaplan-Meier estimate which represents the complier strata if they undergo screening (those assigned to screening who underwent screening). The blue line is the Kaplan-Meier estimate representing those who were invited but did not undergo screening, all of whom are never takers. The red line is the Kaplan-Meier estimate based on those who were not invited to screen, all of whom did not screen, which is a mix of compliers and never takers. The estimate of incidence in an unscreened complier is given by the black line. It is obtained using the method proposed by Baker, which weights the red and blue lines. In particular, it is 1.59 times the red line minus 0.59 times the blue line. We obtained identical results (e.g. to third decimal) using the method which weights observations according to Table 6 and then applies a weighted Kaplan-Meier estimator.

The purpose of our work was to obtain an estimate of the hazard ratio in the compliers, or in other words, an adherence adjusted estimate of the hazard ratio. The ITT estimator was 0.80 (95%CI:0.70–0.92). We applied the principal stratification weights and arrived at a hazard ratio of 0.71 (95% CI: 0.55 – 0.90) as reported in Table 7. The other PSW estimator yielded 0.72 (95% CI: 0.50 – 0.99). In those aged 50–54 the effect of screening is especially strong. Whereas the effect of being invited to screen decreases the hazard of colorectal cancer by 32% (95%CI: 7% to 53%), according to our estimator, the effect of screening in a complier is a 50% (95%CI: 15% to 70%) reduction.

Figure 2 shows that relative to compliers who did not undergo screening, the effect of screening was to increase the incidence of colorectal cancer in the first year. This is to be expected as the cancers are detected earlier because of screening. The reduction in the hazard occurs after the initial phase. Clearly there is a lack of proportionality of hazards due to a crossing of the hazards before 1 year. For this reason we report the ITT and PSW estimators partitioning follow-up into these two time periods in Table 7. Using the PSW estimator we estimate that, whereas screening increases the incidence of diagnosis of colorectal cancer by a factor of 2.64 (95% CI: 1.50 to 4.67)in the first year, it decreases incidence by 50% (95%CI: 0.37 to 0.68) after one year.

5. Discussion

The purpose of our work was to derive an estimator of the hazard ratio in the compliers, analogous to the Complier Average Causal effect (CACE). Other names for this estimator are the adherence adjusted effect and Per Protocol effect (not to be confused with the estimand of the Per Protocol estimator). In simulations we demonstrated that the estimators we propose are unbiased for estimation of the treatment effect, unlike the ITT, Per Protocol and As Treated estimators.

We contend that the CACE and our estimator are patient centered because they deliver a consistent estimator of the effect of the exposure. On the other hand, the ITT delivers a consistent estimator of the effect of assignment to the exposure. The ITT is policy centered and is of important to policy makers. A patient making a decision about a treatment needs to know the CACE.

The reporting of CACE is becoming more common. In addition to the publication of the colorectal cancer screening study of our example, Holme et al. (2014), there have been two other randomized studies reporting CACEs in high tier journals in the past year; Halpern et al. (2015), Perkins et al. (2015). It is important that the limitations of the CACE be repeated. One limitation is that the population for which the CACE and our hazard ratio estimator is applicable is latent; the population of individuals who would have received treatment if randomized to treatment, and not recieve treatment if randomized to control. This population cannot be explicitly identified in any randomized study, although we can report descriptive statistics for this population (Appendix A).

The effect in the latent strata of compliers may be different than in the whole population. For instance, they may be less motivated to obtain the treatment than the strata of always takers because of knowledge that they are less at risk. Similarly, they may be more motivated to obtain the treatment than the strata of never takers because of knowledge that they are more at risk. For example, the strata of compliers in the NORCCAP may contain a high proportion of individuals with knowledge that they are more at risk for colorectal cancer such as family history.

Our approach for using principal stratification to estimate a parameter (e.g. hazard ratio) involves an unmixing of observed distributions to obtain an estimate of the distribution on the treated and untreated compliers. These methods can be implemented using what we call principal stratification weights which can take one of up to four unique values, two of which may be negative. In our experience, negatively valued weights do not occur in any weighting method, with the exception of the method proposed by Abadie (2003) to estimate treatment effects using instrumental variables in the presence of nonlinearity and treatment effect heterogeneity explained by covariates. Not surprisingly, most contemporary statistical software (e.g. R, Stata, SAS) do not facilitate use of negative weights.

We have proposed two approaches for using principal stratification to estimating the hazard ratio for the treatment effect. They differ in the order in which right censoring is taken into account, and therefore the assumptions needed. The easier to implement approach involves implementing methods for right censored data using weighting by prinicipal stratification

weights. This approach assumes that censoring is non-informative conditional on the principal strata. The other approach has the first step of estimating the survival curve (and censoring curve) in the treated and untreated compliers using a weighting of Kaplan-Meier curves, following by an approach of mapping the survival curves (and censoring curves) to a hazard ratio. Both approaches appear to be consistent in simulations. More work is needed to investigate their relative efficiency. A future direction is to adapt the weighting approach at the level of the risk set and to evaluate its relative efficiency.

A limitation of the hazard ratio estimation we propose is that it involves maximizing over a surface which may have more than one local extrema. Our estimator of the hazard ratio is the point that maximizes not the actual likelihood function for the compliers which is not observable but an approximation to this likelihood. While the actual likelihood function is concave, the version of it created using principal stratification weights is not guaranteed to be so. We hypothesize that there is one solution in large samples.

This paper centers around the estimation of the hazard ratio. The hazard ratio is the favored approach for reporting a summary of the treatment effect arising from trials or other studies with survival or other time-to-event outcomes. It has received attention as a causal estimand from Stukel et al. (2007), MacKenzie et al. (2014), Tchetgen Tchetgen et al. (2015) and Wan et al. (2015). The hazard ratio has to be interpreted with the usual cautions. In particular, if there is lack of proportionality of hazards, the hazard ratio is interpreted as an ”average hazard ratio”. The average hazard ratio depends on the follow-up in a study. For example, in our data example the hazard ratio in the first year is inverted compared to the average hazard ratio over the first 10 years. To properly interpret an estimate of a causal hazard ratio the time-frame of the study, or rather the follow-up (censoring) distribution must also be supplied.

Some statisticians and epidemiologists have questioned the causal interpretation of the hazard (Aalen et al. (2015), Hernan et al. (2004)). It is a fact that conditioning on the partial outcome, $T >= t$ induces an association between variables that affect the outcome such as an exposure and a covariate even if they are independent at baseline. In other words, despite the fact that randomization achieves balance in all covariates at baseline, there will not be balance among subjects at risk once follow-up begins and events are accruing (unless the exposure being randomized is ineffective). In the language of instrumental variables, a variable which meets the definition of an instrument at baseline fails to meet that definition at any subsequent point in follow-up. However, there is no logical fallacy in saying the causal effect of a treatment is to alter the hazard in a constant proportional manner.

The model we propose to estimate in this paper is the Cox causal model for the effect of an exposure in the strata of compliers. It is worth emphasizing that this is a marginal model in that it is not conditioning on unmeasured covariates. While there may very well be unmeasured covariates that affect the distribution of the time-to-event (i.e. the hazard) our model supposes that when they are integrated over, a Cox model for the exposure results. The model for the joint effect of the exposure and the unmeasured covariates cannot be written as a proportional hazards model or any particularly convenient form although

MacKenzie et al. (2014) (Appendix A) specify a hybrid multiplicative and additive hazards model for which the marginal distribution of the exposure is a Cox model.

Acknowledgments

This work was partially supported by NIH Grant 1RC4MH092717-01, and through a Patient-Centered Outcomes Research Institute (PCORI) Award ME-1503-28261. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. Magnus Løberg's research was supported by the Norwegian Research Council Grant 231910. The authors have no conflicts of interest to report.

References

Aalen O, Cook R, Roysland K. Does cox analysis of a randomized survival study yield a causal treatment effect? Lifetime Data Analysis. 2015; 113(2):231–63.

Abadie A. Semiparametric instrumental variable estimation of treatment response model. Journal of Econometrics. 2003; 113(2):231–261.

Baker S. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. Journal of the American Statistical Association. 1998; 93:929–934.

Cai B, Small D, TRH. Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. Statistics in Medicine. 2011; 30(15):1809–24. [PubMed: 21495062]

Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. Journal of the Royal Statistical Society: Series B. 2007; 69:565–588.

Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika. 1999; 86(2):365–379.

Halpern S, French B, Small D, Saulsgiver K, Harhay M, Audrain-McGovern J, Loewenstein G, Brennan T, Asch D, KGV. Randomized trial of four financial-incentive programs for smoking cessation. New England Journal of Medicine. 2015; 372(22):2108–17. [PubMed: 25970009]

Hernan M, Hernandez-Diaz S, Robins J. A structural approach to selection bias. Epidemiology. 2004; 15(5):615–25. [PubMed: 15308962]

Hernan M, Hernandez-Diaz S, Robins J. Randomized trials as observational studies. Annals of Internal Medicine. 2013; 159(8):560–2. [PubMed: 24018844]

Holme O, Løberg M, Kalager M, Bretthauer M, Hernn M, Aas E, Eide T, Skovlund E, Schneede J, Tveit K, Hoff G. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. JAMA. 2014; 312(6):606–15. [PubMed: 25117129]

Johnston K, Gustafson P, Levy A, Grootendorst P. Use of ivs in the analysis of generalized linear models in the presence of omitted confounding with applications to epidemiological research. Statistics in Medicine. 2008; 27:1539–1556. [PubMed: 17847052]

Li J, Fine J, Brookhart A. Instrumental variable additive hazards models. Biometrics. 2015; 71(1):122–130. [PubMed: 25298257]

Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics. 2003; 59:100–105. [PubMed: 12762446]

MacKenzie T, Tosteson T, Morden N, Stukel T, OMalley A. Using instrumental variables to estimate a cox's proportional hazards regression subject to additive confounding. Health Services and Outcomes Research Methodology. 2014; 14:54–68. [PubMed: 25506259]

Nie H, Cheng J, Small D. Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. Biometrics. 2011; 67(4):1397–405. [PubMed: 21385167]

Perkins G, Lall R, Quinn T, Deakin C, Cooke M, Horton J, Lamb S, Slowther AM, Woollard M, Carson A, Smyth M, Whitfield R, Williams A, Pocock H, Black J, Wright J, Han K, Gates S. Mechanical versus manual chest compression for out-of-hospital cardiac arrest (paramedic): a pragmatic, cluster randomised controlled trial. The Lancet. 2015; 385:947–55.

Rassen J, Schneeweiss S, Glynn R, Mittleman M, Brookhart M. Iv analysis for estimation of treatment effects with dichotomous outcomes. Am J Epidemiology. 2009; 169(3):273–284.

Shrier I, Kaufman J, Platt R, Steele R. Principal stratification: a broader vision. International Journal of Biostatistics. 2013; 9(2):307–13. [PubMed: 24402884]

Stukel T, Fisher E, Wennberg D, Alter D, Gottlieb D, Vermeulen M. Analysis of observational studies in the presence of treatment selection bias: Effects of invasive cardiac management on ami survival using propensity score and instrumental variable methods. JAMA. 2007; 297(3):278–85. [PubMed: 17227979]

Tchetgen Tchetgen E, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. Epidemiology. 2015; 26(3):402–10. [PubMed: 25692223]

Wan F, Small D, Bekelman J, Mitra N. Bias in estimating the causal hazard ratio when using two-stage instrumental variable methods. Statistics in Medicine. 2015; 34(14):2235–65. [PubMed: 25800789]

A. Baseline Characteristics of the Compliers

The same reasoning above showing how to use weighted subtraction of distributions to estimate the distribution of potential outcomes on the complier strata can be used to estimate the distribution of a baseline characteristic for the complier strata. Let W be a baseline characteristic. Since the conditional distribution of W given $R = X = 1$ is a mixture of W conditional on $(R = 1)\&(PS = Co)$ and $(R = 1)\&(PS = AT)$ with weights $\frac{p_{Co}}{p_{Co} + p_{AT}}$. The distribution of $W|(R = 1)\&(PS = AT)$ equals $W|(R = 0)\&(PS = AT)$ due to randomization (Levy et al, 2004), and the latter equals $W|(R = 0)\&(X = 1)$. Therefore

$$W \mid (R = 1, PS = Co) \sim \frac{p_{Co} + p_{AT}}{p_{Co}} (W \mid R = 1, X = 1) - \frac{p_{AT}}{p_{Co}} (W \mid R = 0, X = 1) . \quad (9)$$

Similarly the conditional distribution of the baseline characteristic W on the compliers randomized to control equals

$$W \mid (R = 0, PS = Co) \sim \frac{p_{Co} + p_{NT}}{p_{Co}} (W \mid R = 0, X = 0) - \frac{p_{NT}}{p_{Co}} (W \mid R = 1, X = 0) . \quad (10)$$

Thus one can report descriptive statistics such as the mean, median, and standard deviation for the baseline characteristics on the compliers by randomization arm. Alternatively one can weight the statistics calculated for each arm by the proportions in each randomization arm to get an overall baseline statistic for the entire sample of compliers.

B. PS2: Accounting for Censoring First

This appendix addresses estimation of the hazard ratio based on using estimators of $T^0(0)|PS = Co$ and $T^0(1)|PS = Co$ which could be done using for instance the approach of Baker (1998). The hazard ratio equals $\ln \frac{\Pr[T^0(1) \geq t \mid PS = Co]}{\Pr[T^0(0) \geq t \mid}$. It is not clear how to proceed, and if it is more efficient to incorporate information about the censoring distributions. As an example of the possible importance of the censoring distributions, consider the setting in which we had independent and identically distributed samples from $\{X, T^0(X), C(X)\}|PS = Co$. In this case the Kaplan-Meier estimators of $T^0(0)|PS = Co$ and $T^0(1)|PS = Co$ would not be sufficient statistics for the hazard ratio, but would need to be combined with Kaplan-Meier estimators of the censoring distributions $C(0)|PS = Co$ and $C(1)|PS = Co$ in order to be sufficient for the maximum partial likelihood estimator (MPLE) of the hazard ratio; that is, the MPLE is a functional of the Kaplan-Meier estimators (MacKenzie et al, 2005). In the simulations we conducted we considered an estimator of the patient centered hazard ratio using this functional of the estimators of $T^0(0)|PS = Co$, $T^0(1)|PS = Co$, $C(0)|PS = Co$ and $C(1)|PS = Co$.

C. Incidence in the Compliers

In this section we derive estimators of the incidence, a metric commonly reported in epidemiology, for the untreated and treated compliers. Under the assumption of a constant hazard function (i.e. exponentially distributed survival) the single hazard parameter can be estimated based on a sample from that distribution by taking the quotient of the number of events (sum of the status indicators) divided by the total follow-up time (sum of follow-up times).

Using the principal stratification weights (2.6), the estimator of the incidence for an untreated complier is

$$\frac{(1+\frac{p_{NT}}{p_{Co}})n_{00}^{-1}\sum_{R_i=0,X_i=0}\Delta_i - \frac{p_{NT}}{p_{Co}}n_{10}^{-1}\sum_{R_i=1,X_i=0}\Delta_i}{(1+\frac{p_{NT}}{p_{Co}})n_{00}^{-1}\sum_{R_i=0,X_i=0}T_i - \frac{p_{NT}}{p_{Co}}n_{10}^{-1}\sum_{R_i=1,X_i=0}T_i} \quad (11)$$

and is

$$\frac{(1+\frac{p_{AT}}{p_{Co}})n_{11}^{-1}\sum_{R_i=1,X_i=1}\Delta_i - \frac{p_{AT}}{p_{Co}}n_{01}^{-1}\sum_{R_i=0,X_i=1}\Delta_i}{(1+\frac{p_{AT}}{p_{Co}})n_{11}^{-1}\sum_{R_i=1,X_i=1}T_i - \frac{p_{AT}}{p_{Co}}n_{01}^{-1}\sum_{R_i=0,X_i=1}T_i} \quad (12)$$

for a treated complier.

D. Code in R for Cox Hazard Ratio Estimation with PSW

```
library(rootSolve)
Cox.2 <- function(T, St, X, Wt) {
    # Maximum Partial Likelihood Estimation with Weights
    # Allows Negative Weights
    miss <- is.na(T + St + X + Wt)
    c(T <- T[!miss], St <- St[!miss], X <- X[!miss], Wt <- Wt[!miss])
    ord <- order(-T)
    c(s.X <- X[ord], s.T <- T[ord], s.St <- St[ord], s.Wt <- Wt[ord])
    f <- function(beta) {
      Denom <- cumsum(s.Wt*exp(beta*s.X))
      Numer.1 <- cumsum(s.Wt*s.X*exp(beta*s.X))
      sum(s.St*s.Wt*(s.X - Numer.1 / Denom), na.rm=T)
      }
    deriv.f <- function(beta) {
      Denom <- cumsum(s.Wt*exp(beta*s.X))
      Numer.1 <- cumsum(s.Wt*s.X*exp(beta*s.X))
      Numer.2 <- cumsum(s.Wt*s.X^2*exp(beta*s.X))
      -sum(s.St*s.Wt* (Numer.2 / Denom - (Numer.1/Denom)^2), na.rm=T)
      }
    est <- multiroot(f, start=0, jacfunc=deriv.f)$root
    }
PS.Weights <- function(X, R) {
    est.P.AT = sum(X==1 & R==0) / sum(R==0)
    est.P.NT = sum(X==0 & R==1) / sum(R==1)
    est.P.Co = 1 - est.P.AT - est.P.NT
PS.Weights <- function(X, R) {
    est.P.AT = sum(X==1 & R==0) / sum(R==0)
    est.P.NT = sum(X==0 & R==1) / sum(R==1)
    est.P.Co = 1 - est.P.AT - est.P.NT
    Wt <- c()
    if (R == 1 & X == 1) Wt <- (1 + est.P.AT/est.P.Co) / sum(R==1& X==1)
    if (R == 1 & X == 0) Wt <- -(est.P.NT/est.P.Co) / sum(R==1 & X==0)
    if (R == 0 & X == 0) Wt <- (1 + est.P.NT/est.P.Co) / sum(R==0 & X==0)
    if (R == 0 & X == 1) Wt <- -(est.P.AT/est.P.Co) / sum(R==0 & X==1)
    Wt
    }
PSW.Cox <- function(T, St, X, R) {
    # T: time, St:status, X: Exposure,0,1
    # R: Instrument (e.g. randomization),0, 1
    # Delivers Complier Hazard Ratio (log) X=1 vs X=0
    Wt <- PS.Weights(X, R)
    o <- Cox.2(T, St, X, Wt)
```

```
list(coef=o)
}
```

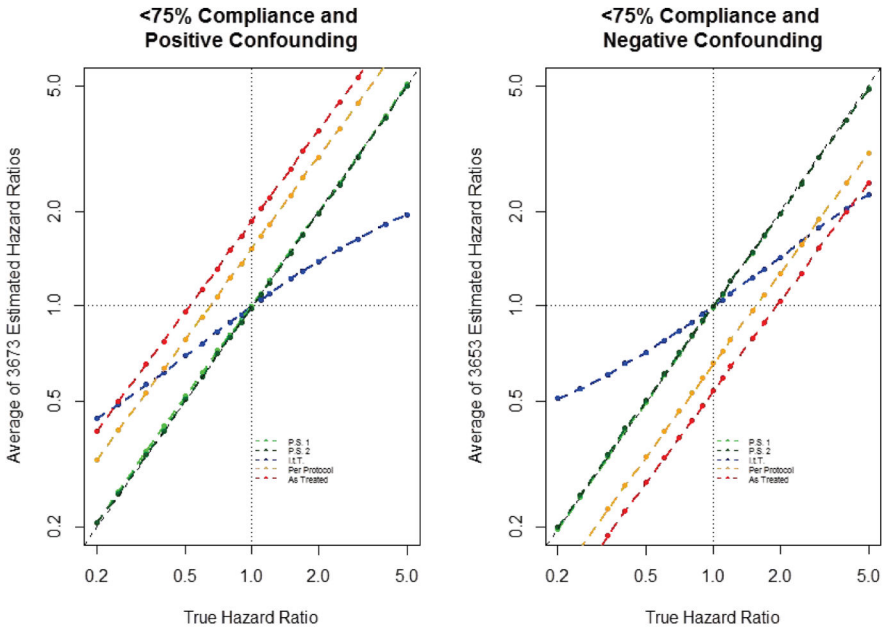

Figure 1.

Simulation Results: A plot of average estimated hazard ratio (geometric mean) versus actual hazard ratio for the ITT, Per-Protocol, As Treated and our proposals for a hazard ratio estimator based upon principal stratification. PS1 corresponds to the PSW estimator described in the text, while PS2 corresponds to the estimator described in Appendix B. An unbiased estimator is characterized by its proximity to the line of identity.

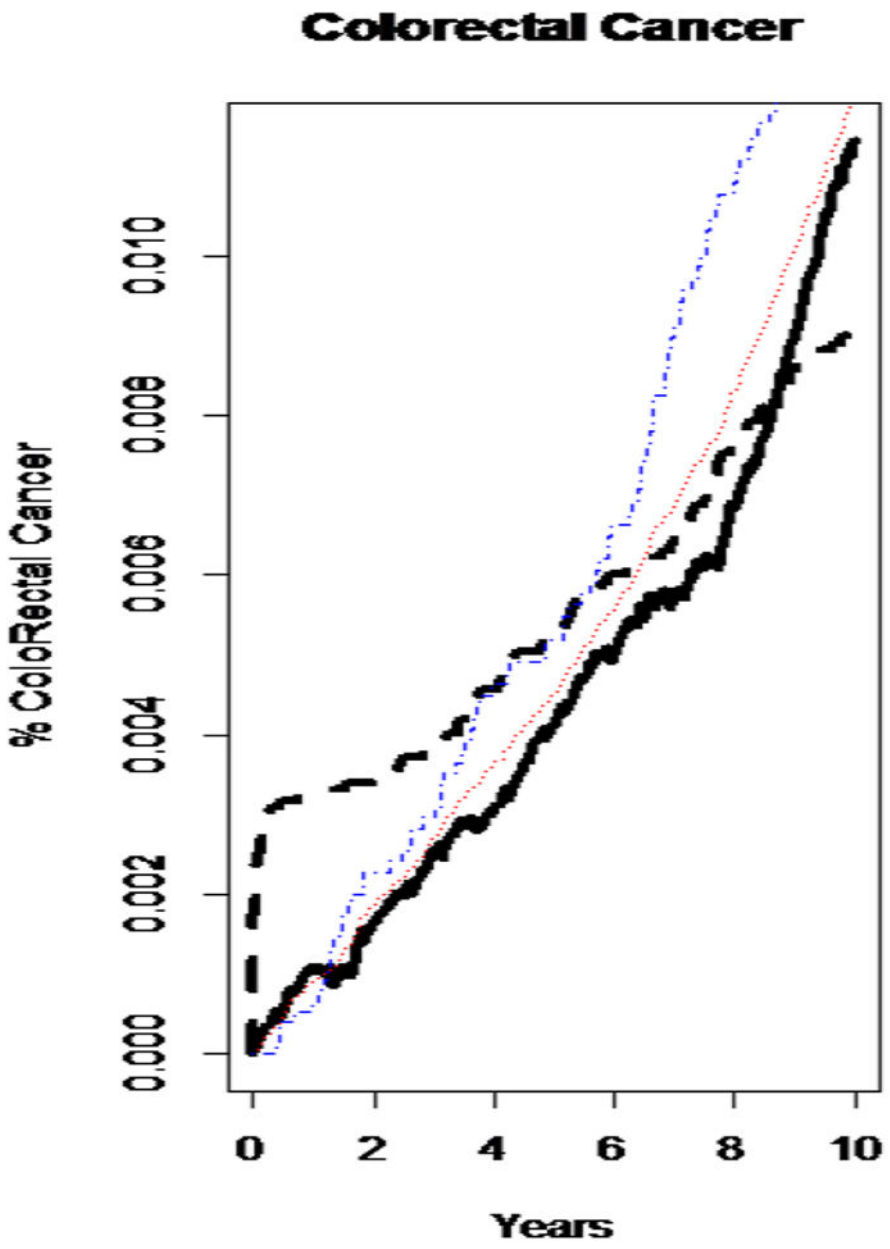

Figure 2.

Cumulative incidence of CRC up to 10 years after invitation to screen. The dashed thick curve represents CRC incidence were a complier to receive screening, whereas the solid thick curve represents CRC incidence were a complier to to receive screening. The solid curve is composed as 1.59 times the red dotted curve (CRC incidence in the control arm) minus 0.59 times the blue dashed curve (CRC incidence in subjects invited to screen but who declined).

Table 1

Mappings from Domain of the Instrument, R, to Domain of the Instrument, X, and the Corresponding Principal Strata Name

Mapping	Name
$(0, 1) \mapsto (0, 0)$	Never Takers
$(0, 1) \mapsto (0, 1)$	Compliers
$(0, 1) \mapsto (1, 0)$	Defiers
$(0, 1) \mapsto (1, 1)$	Always Takers

Table 2

Determination of principal strata according to the values of the instrument, *R*, and exposure, *X*

R	X	Principal Strata
0	0	Compliers or Never Takers
0	1	Always Takers or Defiers
1	0	Never Takers or Defiers
1	1	Compliers or Always Takers

Table 3

Principal stratification weights (PSW) by assignment and exposure where $d = n_{00}n_{11} - n_{01}n_{10}$.

Exposure	Assignment R = 0	Assignment R = 1
X=0	$\{1 + \frac{\hat{p}_{NT}}{\hat{p}_{C0}}\}n_0 \cdot n_{00}^{-1} = n_0 \cdot n_1 \cdot d^{-1}$	$-\frac{\hat{p}_{NT}}{\hat{p}_{C0}}n_0 \cdot n_{10}^{-1} = -n_0^2 \cdot d^{-1}$
X=1	$-\frac{\hat{p}_{AT}}{\hat{p}_{C0}}n_1 \cdot n_{01}^{-1} = -n_1^2 \cdot d^{-1}$	$\{1 + \frac{\hat{p}_{AT}}{\hat{p}_{C0}}\}n_1 \cdot n_{11}^{-1} = n_0 \cdot n_1 \cdot d^{-1}$

Table 4

Simulation Variables and Their Random Generation

Name	Symbol	Data Generation
Latent Confounder	Z	Standard Normal
Principal Strata	PS	Ordinal Logistic(·; 5Z, p_{NT}, p_{AT}, p_{Co})
Instrument	R	Bernoulli(1/2)
Exposure	X	$R \cdot 1_{PS=Co} + 1_{PS=AT}$
Potential time-to-event if untreated	$T^0(0)$	Exponential(λ_{PS})
Potential time-to-event if treated	$T^0(1)$	Exponential(exp($\beta X + \phi Z$)λ_{PS})
Censoring time	C	Uniform(0, ρ)
Follow-up if untreated	$T(0)$	min($T^0(0)$, C)
Follow-up if treated	$T(1)$	min($T^0(1)$, C)
Status indicator if untreated	$\Delta(0)$	$1_{T^0(0)<C}$
Status indicator if treated	$\Delta(1)$	$1_{T^0(1)<C}$

Table 5

Ranges of the Parameters Used in the Simulation

Name	Symbol	Range
Hazard Ratio	$\exp(\beta)$	0.2, 0.3, 0.5, 0.7, 0.8, 1, 1.2, 1.5, 2, 3, 5
No. of Events	N	50, 100, 150, 200
Censoring %	f_C	Uniform(0,90)
Pr[PS=AT]	p_{AT}	Uniform(0, 0.4)
Pr[PS=NT]	p_{NT}	Uniform(0, 0.4)
Pr[PS=Co]	p_{C0}	$1 - p_{NT} - p_{AT}$
ln Hazard Ratio T^0 vs Latent Confounder Z	ϕ	Uniform(ln(1/3), ln(3))
Baseline Hazard in Always Takers	λ_{AT}	exp(Uniform(log(1/3),log(3))
Baseline Hazard in Never Takers	λ_{NT}	exp(Uniform(log(1/3),log(3))
Baseline Hazard in Compliers	λ_{Co}	exp(Uniform(log(1/3),log(3))

Table 6

Principal stratification weights and number of subjects for the NORCCAP randomized trial according to randomized assignment and exposure recieved. The weight in the lower right cell is not applicable because there were no individuals who were randomized not to recieve a screening invitation who received screening

	Randomization	
Screening Recieved	**No Invitation N=78,220**	**Invitation to Screen N=20,572**
No screening N=85,837	1.59 · 85837/78220 = **1.74** N=78,220	−0.59 · 85837/7617 = **−6.64** N=7,617
Screening N=12,955	Not applicable N=0	1.00 · 12955/12955 = **1.00** N=12,955

Table 7

Hazard Ratio Estimates with 95% Confidence Intervals

		Principal Stratification	
	Intention to Treat	**Version One**	**Version Two**
Overall	0.80 (0.70–0.92)	0.71 (0.55–0.90)	0.72 (0.52–0.99)
Ages 50–54	0.68 (0.47–0.93)	0.50 (0.30–0.85)	0.49 (0.21–1.16)
Ages 55–64	0.86 (0.72–1.02)	0.77 (0.59–1.02)	0.76 (0.54–1.09)
Year 1	2.56 (1.77–3.70)	2.64 (1.50–4.67)	3.13 (1.19–8.26)
Years 2 and up	0.74 (0.62–0.87)	0.50 (0.37–0.68)	0.55 (0.41–0.75)